RUSSELL K. SCHUTT

INVESTIGATING THE SOCIAL WORLD

The Process and Practice of Research

TENTH EDITION

Investigating the Social World

Tenth Edition

To Julia Ellen Schutt

Sara Miller McCune founded SAGE Publishing in 1965 to support the dissemination of usable knowledge and educate a global community. SAGE publishes more than 1000 journals and over 800 new books each year, spanning a wide range of subject areas. Our growing selection of library products includes archives, data, case studies and video. SAGE remains majority owned by our founder and after her lifetime will become owned by a charitable trust that secures the company's continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

Investigating the Social World

The Process and Practice of Research

Tenth Edition

Russell K. Schutt

University of Massachusetts Boston

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London, EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
18 Cross Street #10-10/11/12
China Square Central
Singapore 048423

Copyright © 2022 by SAGE Publications, Inc.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third-party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.

Printed in Canada

ISBN: 978-1-0718-1705-6

Library of Congress Control Number: 2022913741

Acquisitions Editor: Helen Salmon

Content Development Editor: Darcy Scelsi

Production Editor: Veronica Stapleton Hooper

Copy Editor: Karin Rathert

Typesetter: diacriTech

Cover Designer: Scott Van Atta

Marketing Manager: Victoria Velasquez

This book is printed on acid-free paper.

22 23 24 25 26 10 9 8 7 6 5 4 3 2 1

BRIEF CONTENTS

About the Author xix

Preface xxi

Acknowledgments xxix

SECTION I FOUNDATIONS FOR SOCIAL RESEARCH 1

Chapter 1 Science, Society, and Social Research 2

Chapter 2 The Process and Problems of Social Research 31

Chapter 3 Research Ethics and Research Proposals 75

SECTION II FUNDAMENTALS OF SOCIAL RESEARCH 109

Chapter 4 Conceptualization and Measurement 110

Chapter 5 Sampling and Generalizability 150

Chapter 6 Research Design and Causation 193

SECTION III BASIC SOCIAL RESEARCH DESIGNS 227

Chapter 7 Experiments 228

Chapter 8 Survey Research 260

Chapter 9 Quantitative Data Analysis 301

Chapter 10 Qualitative Methods 349

Chapter 11 Qualitative Data Analysis 387

SECTION IV COMPLEX SOCIAL RESEARCH DESIGNS 429

Chapter 12 Mixed Methods 430

Chapter 13 Evaluation and Policy Research 453

Chapter 14 Using Secondary Data and Big Data 488

Chapter 15 Historical and Comparative Methods and Content Analysis 521

Chapter 16 Summarizing and Reporting Research 559

Appendix A: Questions to Ask About a Research Article 592

Appendix B: Table of Random Numbers 594

Glossary 599

Bibliography 616

Index 662

DETAILED CONTENTS

About the Author xix

Preface xxi

Acknowledgments xxix

SECTION I FOUNDATIONS FOR SOCIAL RESEARCH 1

Chapter 1 Science, Society, and Social Research 2

Learning About the Social World 4
Motives for Social Research 7
Avoiding Errors in Reasoning About the Social World 8
Observing 9
Generalizing 11
Reasoning 12
Reevaluating 12
Science and Social Science 14
The Scientific Approach 15
Pseudoscience or Science 17
Types of Social Research 18
Descriptive Research 18
Exploratory Research 19
Explanatory Research 20
Evaluation Research 21
Strengths and Limitations of Social Research 22
Alternative Research Orientations 23
Quantitative and/or Qualitative Methods 23
Philosophical Perspectives 24
Basic Science or Applied Research 24
The Role of Values 24
Conclusions 25
Key Terms 27
Highlights 27
Discussion Questions 29
SPSS Exercises 29

Chapter 2 The Process and Problems of Social Research 31

Social Research Questions 33
Identifying Social Research Questions 34
Refining Social Research Questions 34
Evaluating Social Research Questions 35
Feasibility 35

Social Importance 35
Scientific Relevance 36
Social Theories 37
Scientific Paradigms 39
Searching for Prior Research 40
Searching the Literature 40
Searching the Web 43
Reviewing Prior Research 45
Single-Article Reviews: Aggressive Policing and Educational Performance of Minority Youth 47
Integrated Literature Reviews: When Does Arrest Matter? 49
Systematic Literature Reviews: Second Responder Programs and Repeat Family Abuse Incidents 51
Social Research Strategies 56
Explanatory Research 58
Deductive Research 58
Body-Worn Cameras and the Research Circle 60
Inductive Research 62
An Inductive Approach to Explaining Support for Black Lives Matter 63
Exploratory Research 64
Black Youths' Strategies for Navigating Police Contact 64
Descriptive Research 65
Social Research Standards 66
Measurement Validity 67
Generalizability 68
Causal Validity 69
Authenticity 70
Conclusions 70
Key Terms 71
Highlights 71
Discussion Questions 73
SPSS Exercises 73

Chapter 3 Research Ethics and Research Proposals 75

Historical Background 78
Ethical Principles 81
Achievement of Valid Results 81
Honesty and Openness 83
Protection of Research Participants 85
Avoid Harming Research Participants 85
Obtain Informed Consent 89
Avoid Deception in Research, Except in Limited Circumstances 90
Maintain Privacy and Confidentiality 92
Consider Uses of Research So That Benefits Outweigh Risks 92
Informed Consent Forms 94
The Institutional Review Board 98
Social Research Proposals 100
Case Study: Evaluating a Public Health Program 103

Conclusions 105
Key Terms 106
Highlights 106
Discussion Questions 108
SPSS Exercises 108

SECTION II FUNDAMENTALS OF SOCIAL RESEARCH 109

Chapter 4 Conceptualization and Measurement 110

Concepts 112
Conceptualization in Practice 113
Tolerance 113
Poverty 113
From Concepts to Indicators 117
Operationalizing the Concept of Race 118
Operationalizing Perceived Threat From Immigrants 120
The Inductive Approach to Conceptualization 122
Measurement 122
Constructing Questions 123
Making Observations 126
Collecting Unobtrusive Measures 127
Physical Trace Evidence 127
Using Available Data 127
Coding Content 128
Taking Pictures 128
Combining Measurement Operations 129
Levels of Measurement 130
Nominal Level of Measurement 131
Ordinal Level of Measurement 132
Interval Level of Measurement 133
Ratio Level of Measurement 133
The Special Case of Dichotomies 135
Comparison of Levels of Measurement 135
Evaluating Measures 136
Measurement Validity 136
Face Validity 137
Content Validity 138
Criterion Validity 138
Construct Validity 138
Measurement Reliability 139
Multiple Times: Test–Retest and Alternate Forms 139
Multiple Indicators: Interitem and Split-Half 140
Multiple Observers: Interobserver and Intercoder 140
Ways to Improve Reliability and Validity 142
Conclusions 143
Key Terms 144
Highlights 144

Discussion Questions 147
SPSS Exercises 149

Chapter 5 Sampling and Generalizability 150

Sample Planning 152
The Purpose of Sampling 152
Define Sample Components and the Population 152
Sample Generalizability 154
Assess the Diversity of the Population 156
Consider a Census 157
Probability and Representativeness 158
Probability Sampling Methods 164
Simple Random Sampling 165
Systematic Random Sampling 166
Stratified Random Sampling 167
Multistage Cluster Sampling 172
Probability Sampling Methods Compared 173
Nonprobability Sampling Methods 174
Availability (Convenience) Sampling 175
Quota Sampling 176
Purposive Sampling 178
Snowball Sampling 179
Respondent-Driven Sampling 179
Lessons About Sample Quality 179
Generalizability With Web-Based Samples 181
Generalizability in Qualitative Research 182
Sampling Distributions 183
Estimating Sampling Error 184
Sample Size Considerations 186
Conclusions 187
Key Terms 187
Highlights 188
Discussion Questions 191
SPSS Exercises 191

Chapter 6 Research Design and Causation 193

Units of Analysis 196
Individual and Group 196
The Ecological Fallacy and Reductionism 198
Cross-Sectional and Longitudinal Designs 200
Cross-Sectional Designs 200
Longitudinal Designs 202
Nomothetic or Idiographic Causal Explanations 207
Nomothetic Causal Explanations 208
Idiographic Causal Explanations 209
Criteria for Nomothetic Causal Explanations 210
Association 212

Time Order 213
Experimental Designs 213
Nonexperimental Designs 213
Nonspuriousness 214
Randomization 216
Statistical Control 216
Cautions for Nomothetic Causal Explanations 218
Mechanism 218
Context 219
Comparing Research Designs 220
Conclusions 222
Key Terms 222
Highlights 223
Discussion Questions 225
SPSS Exercises 226

SECTION III BASIC SOCIAL RESEARCH DESIGNS 227

Chapter 7 Experiments 228

True Experiments 230
Experimental and Comparison Groups 230
Pretest and Posttest Measures 232
Randomization 233
Limitations of True Experimental Designs 234
Summary: Causality in True Experiments 235
Association Between the Hypothesized Independent and Dependent Variables 236
Time Order of Effects of One Variable on the Others 236
Nonspurious Relationships Between Variables 236
Mechanism That Creates the Causal Effect 236
Context in Which Change Occurs 236
Quasi-Experiments 238
Nonequivalent Control Group Designs 238
Aggregate Matching 238
Individual Matching 239
Before-and-After Designs 240
Summary: Causality in Quasi-Experiments 242
Association Between the Hypothesized Independent and Dependent Variables 242
Time Order of Effects of One Variable on the Others 242
Nonspurious Relationships Between Variables 242
Mechanism That Creates the Causal Effect 243
Context in Which Change Occurs 243
Nonexperimental Ex Post Facto Control Group Designs 244
Causal (Internal) Validity in Experiments 244
Sources of Internal Invalidity Reduced by a Comparison Group 244
Sources of Internal Invalidity Reduced by Randomization 245
Sources of Internal Invalidity That Require Attention While the Experiment Is in Progress 246

Generalizability in Experiments 248
Sample Generalizability 248
Factorial Surveys 249
External Validity 251
Interaction of Testing and Treatment 251
Ethical Issues in Experimental Research 252
Deception 252
Selective Distribution of Benefits 254
Conclusions 255
Key Terms 255
Highlights 256
Discussion Questions 258
SPSS Exercises 259

Chapter 8 Survey Research 260

Survey Research in the Social Sciences 261
Attractions of Survey Research 262
Versatility 262
Efficiency 262
Generalizability 262
The Omnibus Survey 263
Errors in Survey Research 264
Poor Measurement 264
Nonresponse 264
Inadequate Coverage of the Population 264
Sampling Error 265
Writing Survey Questions 265
Avoid Confusing Phrasing 265
Minimize the Risk of Bias 267
Maximize the Utility of Response Categories 267
Avoid Making Either Disagreement or Agreement Disagreeable 269
Minimize Fence-Sitting and Floating 271
Combining Questions in Indexes 272
Designing Questionnaires 275
Build on Existing Instruments 275
Refine and Test Questions 275
Add Interpretive Questions 276
Maintain Consistent Focus 277
Order the Questions 277
Make the Questionnaire Attractive 278
Consider Language and Culture 279
Organizing Surveys 279
Mailed, Self-Administered Surveys 280
Group-Administered Surveys 283
Telephone Surveys 284
Reaching Sample Units 285
Maximizing Response to Phone Surveys 284

In-Person Interviews 286
Maximizing Response to Interviews 287
Web Surveys 288
Mixed-Mode Surveys 290
A Comparison of Survey Designs 291
Ethical Issues in Survey Research 293
Conclusions 294
Key Terms 294
Highlights 295
Discussion Questions 298
SPSS Exercises 299

Chapter 9 Quantitative Data Analysis 301

Introducing Statistics 303
Case Study: The Likelihood of Voting 303
Preparing for Data Analysis 306
Displaying Univariate Distributions 307
Graphs 307
Bar Chart 308
Histograms 308
Frequency Polygon 309
Misuse of Graphs 310
Frequency Distributions 311
Ungrouped Data 312
Grouped Data 312
Combined and Compressed Distributions 314
Summarizing Univariate Distributions 315
Measures of Central Tendency 316
Mode 316
Median 317
Mean 317
Median or Mean? 318
Measures of Variation 320
Range 321
Interquartile Range 322
Variance 322
Standard Deviation 323
Cross-Tabulating Variables 324
Constructing Contingency Tables 324
Graphing Association 327
Describing Association 329
Evaluating Association 330
Controlling for a Third Variable 333
Intervening Variables 333
Extraneous Variables 334
Specification 337
Regression Analysis 338
Analyzing Data Ethically 341

Conclusions 343
Key Terms 343
Highlights 344
Discussion Questions 347
SPSS Exercises 347

Chapter 10 Qualitative Methods 349

Fundamentals of Qualitative Methods 351
History of Qualitative Research 351
Features of Qualitative Research 352
Collection Primarily of Qualitative Rather Than Quantitative Data 352
A Focus on Previously Unstudied Processes and Unanticipated Phenomena 352
Exploratory Research Questions, With a Commitment to Inductive Reasoning 352
Sensitivity to the Subjective Role of the Researcher (Reflexivity) 353
An Orientation to Social Context, to the Interconnections Between Social Phenomena Rather Than to Their Discrete Features 353
A Focus on Human Subjectivity, on the Meanings That Participants Attach to Events and That People Give to Their Lives 353
Use of Idiographic Rather Than Nomothetic Causal Explanation 353
Acceptance—by Some Qualitative Researchers—of a Constructivist Philosophy 354
Adaptive Research Design, in Which the Design Develops as the Research Progress 355
Styles of Qualitative Research 355
The Case Study 355
Ethnography 356
Digital Ethnography 357
Hybrid Ethnography 357
Participant Observation 358
Choosing a Role 358
Covert Observation 358
Overt Observation 360
Overt Participation (Participant Observer) 360
Covert Participation 361
Entering the Field 363
Developing and Maintaining Relationships 363
Sampling People and Events 364
Taking Notes 367
Managing the Personal Dimensions 367
Intensive Interviewing 369
Establishing and Maintaining a Partnership 371
Asking Questions and Recording Answers 372
Interviewing Online 372
Focus Groups 374
Generalizability in Qualitative Research 375
Ethical Issues 376
Voluntary Participation 377
Subject Well-Being 377
Identity Disclosure 378
Confidentiality 379

Appropriate Boundaries 379
Researcher Safety 380
Conclusions 381
Key Terms 382
Highlights 382
Discussion Questions 385
SPSS Exercises 386

Chapter 11 Qualitative Data Analysis 387

Features of Qualitative Data Analysis 389
Qualitative Data Analysis as an Art 390
Qualitative Compared With Quantitative Data Analysis 392
Techniques of Qualitative Data Analysis 393
Documentation 394
Coding, Categorization, and Condensation 395
Examination and Display of Relationships 399
Corroboration and Legitimization of Conclusions 400
Reflection on the Researcher's Role 401
Alternatives in Qualitative Data Analysis 402
Grounded Theory 402
Abductive Analysis 404
Case-Oriented Understanding 404
Conversation Analysis 405
Narrative Analysis 406
Qualitative Comparative Analysis 409
Combining Qualitative Methods 410
Visual Sociology 411
Systematic Observation 414
Participatory Action Research 415
Computer-Assisted Qualitative Data Analysis 418
Ethics in Qualitative Data Analysis 420
Privacy, Confidentiality, and Anonymity 422
Intervention and Advocacy 423
Research Integrity andQquality 423
Ownership of Data and Conclusions 423
Use and Misuse of Results 424
Conclusions 424
Key Terms 425
Highlights 425
Discussion Questions 427
Computer-Assisted Qualitative Data Analysis 428

SECTION IV COMPLEX SOCIAL RESEARCH DESIGNS 429

Chapter 12 Mixed Methods 430

History of Mixed Methods 432
Diagramming Mixed Methods 434

Types of Mixed Methods 436
Integrated Mixed-Methods Designs 436
Embedded Mixed-Methods Designs 438
Staged Mixed-Methods Designs 438
Complex Research Program Mixed-Methods Designs 440
Strengths and Limitations of Mixed Methods 443
Ethics and Mixed Methods 448
Conclusions 449
Key Terms 450
Highlights 450
Discussion Questions 452
SPSS Exercises 452

Chapter 13 Evaluation and Policy Research 453

History of Evaluation Research 456
Evaluation Basics 457
Questions for Evaluation Research 459
Needs Assessment 459
Evaluability Assessment 461
Process Evaluation 463
Impact Analysis 466
Efficiency Analysis 469
Design Decisions 471
Black Box Evaluation or Program Theory 471
Researcher or Stakeholder Orientation 473
Quantitative or Qualitative Methods 475
Simple or Complex Outcomes 476
Groups or Individuals 478
Policy Research 478
Ethics in Evaluation 480
Conclusions 483
Key Terms 483
Highlights 484
Discussion Questions 486
SPSS Exercises 487

Chapter 14 Using Secondary Data and Big Data 488

Secondary Data Analysis 490
Overview 490
Inter-university Consortium for Political and Social Research 492
Types of Data Available From ICPSR 492
Obtaining Data From ICPSR 493
Secondary Data From Government and Other Sources 495
U.S. Census Bureau 496
Integrated Public Use Microdata Series 496
Bureau of Labor Statistics 497

Other Government Sources 497
Other Data Sources 497
Societal Experts Action Network (SEAN) 498
Harvard's Dataverse 499
International Data Sources 499
Qualitative Data Sources 499
Challenges for Secondary Data Analyses 500
Big Data 505
Background 505
Examples of Research Using Big Data 507
Digital Methods 509
Challenges for Using Big Data 511
Ethical Issues With Secondary Data and Big Data 512
Conclusions 516
Key Terms 517
Highlights 517
Discussion Questions 519
SPSS Exercise 520

Chapter 15 Historical and Comparative Methods and Content Analysis 521

Overview of Historical and Comparative Research Methods 522
Historical Social Science Methods 524
Historical Events Research 525
Event-Structure Analysis 526
Oral History 526
Historical Process Research 528
Cautions for Historical Methods 529
Comparative Social Science Methods 531
Comparative Cross-Sectional Research 531
Comparative Historical Research 536
Comparative Case Study Designs 540
Cautions for Comparative Methods 541
Demographic Analysis 543
Content Analysis 545
Identify a Population of Documents or Other Textual Sources 548
Determine the Units of Analysis 548
Select a Sample of Units From the Population 549
Design Coding Procedures for the Variables to Be Measured 549
Develop Appropriate Statistical Analyses 553
Ethical Issues in Historical and Comparative Research and Content Analysis 555
Conclusions 555
Key Terms 556
Highlights 556
Discussion Questions 558
SPSS Exercises 558

Chapter 16 Summarizing and Reporting Research 559

Writing Research 560
Displaying Research 562
Performing Meta-Analyses 565
Case Study: Vaccine Intentions Across 28 Large National Samples 567
Reporting Research 569
Journal Articles 570
Case Study: Does Protest Against Police Violence Matter? 570
Applied Research Reports 572
Findings From Living in Boston During COVID-19: Economic Impact 573
Framing an Applied Report 575
Research Videos 577
Research Posters and Presentations 577
Reporting Quantitative and Qualitative Research 580
Ethics, Politics, and Research Reports 582
Plagiarism 584
Conclusions 586
Key Terms 587
Highlights 587
Discussion Questions 590
SPSS Exercises 590

Appendix A: Questions to Ask About a Research Article 592

Appendix B: Table of Random Numbers 594

Glossary 599

Bibliography 616

Index 662

ABOUT THE AUTHOR

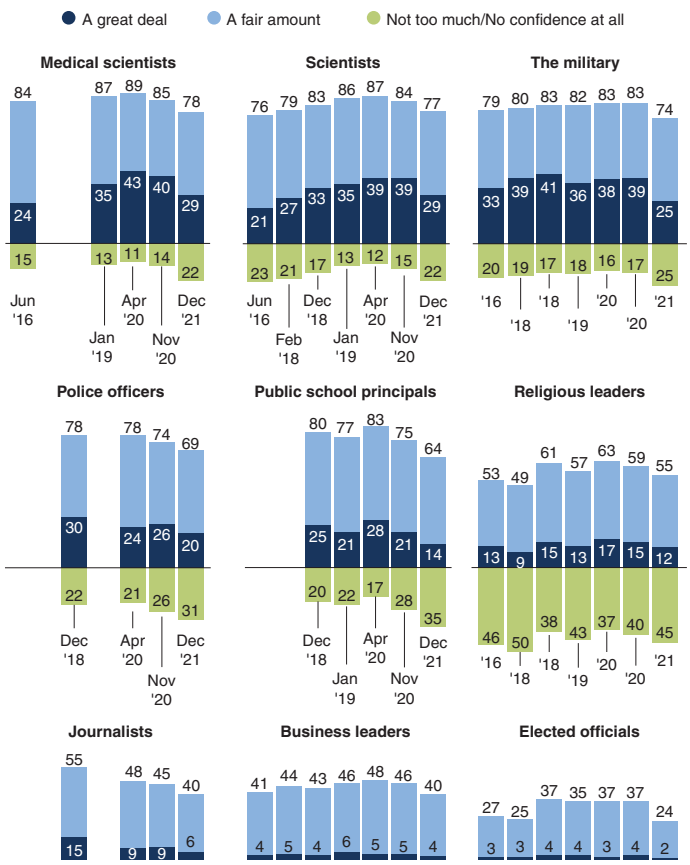
Russell K. Schutt, PhD, is professor of sociology at the University of Massachusetts Boston; Clinical Research Scientist I at the Beth Israel Deaconess Medical Center; and lecturer (part-time) in the Department of Psychiatry, Harvard Medical School. He completed his BA, MA, and PhD degrees at the University of Illinois at Chicago and his postdoctoral fellowship in the Sociology of Social Control Training Program at Yale University. In addition to *Investigating the Social World: The Process and Practice of Research* and adaptations of that text—*Understanding the Social World: Research Methods for the 21st Century*; *Making Sense of the Social World* (with Dan Chambliss); *Research Methods in Psychology* (with Paul G. Nestor); *The Practice of Research in Criminology and Criminal Justice* and *Fundamentals of Research in Criminology and Criminal Justice* (with Ronet Bachman); *The Practice of Research in Social Work* and *Fundamentals of Social Work Research* (with Ray Engel); and *Research Methods in Education* (with Joseph Check)—he is the author of *Homelessness, Housing, and Mental Illness* and *Organization in a Changing Environment*, coeditor of *Social Neuroscience: Brain, Mind, and Society* and of *The Organizational Response to Social Problems*, and coauthor of *Responding to the Homeless: Policy and Practice.* He has authored and coauthored more than 65 peer-reviewed journal articles as well as many book chapters and research reports on homelessness, mental health, service preferences and satisfaction, organizations, and the sociology of law. His recent research has been funded by the National Science Foundation (survey on the social impact of the pandemic in Boston), with collaborators at the Center for Survey Research (UMass Boston) and Northeastern University; the Patient-Centered Outcomes Research Institute (PCORI) at the Beth Israel Deaconess Medical Center with collaborators at the Harvard Medical School and other universities (randomized comparison of two socially oriented interventions to improve community functioning among persons diagnosed with serious mental illness); and the Veterans Health Administration, with colleagues at the University of Massachusetts Medical School and the VA (randomized trial of peer support). His past research has been funded by the National Cancer Institute, the Veterans Health Administration, the National Institute of Mental Health, the Fetzer Institute, and state agencies. Details are available at https://blogs.umb.edu/russellkschutt/.

PREFACE

The COVID-19 pandemic years were tough on confidence in scientists and other established authorities in many countries. Although confidence in medical scientists and scientists in general remains high, you can see in Figure P.1 that confidence in both declined about 10% by December 2021 after having reached a high of 89% and 87%, respectively, at the start of the pandemic. The Pew Research Survey results in Figure P.1 show that confidence in the military, religious leaders, and business leaders also declined about as much, but confidence declined even more in relation to police officers, public school principals, journalists, and elected officials.

FIGURE P.1 ■ Public Confidence in Scientists and Medical Scientists Has Declined Over the Last Year | Pew Research Center

% of U.S. adults who have___of confidence in the following groups to act in the best interests of the public



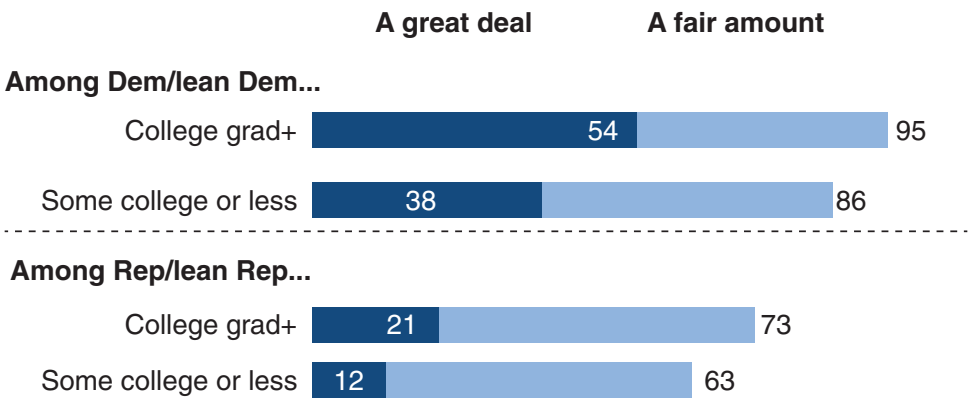
Source: Kennedy, Brian, Alec Tyson, and Cary Funk. "Americans' Trust in Scientists, Other Groups Declines." Pew Research Center, Washington, DC, February 2022. https://www.pewresearch.org/science/2022/02/15/americans-trust-in-scientists-other-groups-declines/.

Are you reassured by the relatively high level of confidence that Americans still have for medical and other scientists, despite the turmoil during the pandemic, or are you disappointed to learn that confidence in science didn't increase? After all, scientists developed remarkably effective new vaccines and health care workers and public officials saved many lives by applying lessons from medical science and from the science of public health. Social science can't tell you whether to be reassured or disappointed, but social science research methods can improve our understanding of the social world by guiding the design of surveys like the ones Pew conducts. With the type of systematic data in Figure P.1, we can begin to see beyond the lmits of our own experience (Kennedy, Tyson, and Funk 2022).

The picture gets more interesting when the Pew researchers consider the educational level and political party preference of the survey respondents. Confidence in medical scientists increased with educational level and preference for the Democratic party. The result of these two trends is shown in Figure P.2 (Kennedy et al. 2022). Among college graduates who identified as preferring the Democratic party, confidence in medical scientists was about 20 percentage points higher than among less-educated persons who preferred the Republican party. Another figure in the report shows that level of confidence in medical scientists only declined among Republicans after the pandemic began, not among Democrats.

FIGURE P.2 ■ Confidence in Medical Scientists by Education and Political Preference

% of U.S adults who have___of confidence in medical scientists to act_in the best interests of the public



Note: Respondents who gave other responses or did not give an answer are not shown.

Source: Kennedy, Brian, Alec Tyson, and Cary Funk. "Americans' Trust in Scientists, Other Groups Declines." Pew Research Center, Washington, DC, February 2022. https://www.pewresearch.org/science/2022/02/15/americans-trust-in-scientists-other-groups-declines/

You now have more information about the underpinnings of confidence in medical scientists. You know such confidence varies with educational level and, since the pandemic began, with political party preference. Is this pattern consistent with your own feelings about medical scientists and those of your friends? Do these findings raise other questions for you? Are younger people more confident in medical scientists? If so, could that be because they tend to be more educated? Might these

patterns vary across the regions of the United States? Do you think similar patterns would occur in other countries?

And what about the methods behind this research? Do you wonder how these findings were obtained? Are you sure the information is trustworthy—that the findings are correct? If you read the methodological appendix in the full report by Brian Kennedy, Alec Tyson, and Cary Funk (2022), you'll learn that these findings come from a survey of 14,497 adults who participate in Pew's ongoing American Trends Panel. But is that all you need to know to assess the trustworthiness of these findings? Do you want to know what questions were asked in the survey and how the adults were selected?

These answers may in turn raise more questions: What is meant in the methodological appendix by the term "stratified random sample"? Is a "response rate" of 65% adequate? What are "respondent weights" and "sampling error"? How good are the questions asked in the Pew survey?

Of course, I have presented only a bit of the evidence about the decline of confidence in medical scientists and others and just a few of the conclusions about variation in confidence between social groups and over time; there are many related issues to consider and important questions to ask about the evidence. My intent now is not to give you the whole picture about recent declines in confidence but to introduce the study of research methods by illustrating how research helps us understand pressing social questions and why we need to learn more about research methods to evaluate the results of such research.

Neither our own limited perceptions nor a few facts gleaned from even reputable sources provide us with a trustworthy basis for understanding the social world. We need systematic methods for investigating our social world that enable us to chart our course through the passions of the moment and to see beyond our own personal experience.

TEACHING AND LEARNING GOALS

If you see the importance of pursuing answers to questions about the social world the way that the Pew Research Center does, you can understand the importance of investigating the social world. One purpose of this book is to introduce you to social science research methods such as those involved in the Pew study of confidence in medical scientists and to show how they improve everyday methods of answering our questions about the social world. Each chapter integrates instruction in research methods with the investigation of interesting aspects of the social world, with many examples in this tenth edition reflecting investigations about the tumultuous social world that we have experienced since 2020. You'll learn about online social connections, police use of force, feelings of tolerance, changes in the homicide rate, the response to disasters, and much more.

Another purpose of this book is to give you the critical skills necessary to evaluate research. Just "doing research" is not enough. Just reading that some conclusions are "based on a research study" is not sufficient. You must learn to ask many questions before concluding that research-based conclusions are appropriate. What did the researchers set out to investigate? How were people selected for study? What information was collected, and how was it analyzed? Throughout this book, you will learn what questions to ask when critiquing a research study and how to evaluate the answers. You can begin to sharpen your critical teeth on the illustrative studies throughout the book.

A third goal of this book is to prepare you to do research. Substantive examples will help you see how methods are used in practice. Exercises at the end of each chapter and on the book study site give you ways to try different methods. A checklist for research proposals will help you plan more ambitious studies. But research methods cannot be learned by rote and applied mechanically. Thus, you will learn the benefits and liabilities of each major approach to research and why employing a combination of them is often preferable. You will come to appreciate why the results of single research studies must always be interpreted within the context of prior research and through the lens of social theory.

ORGANIZATION OF THE BOOK

The way the book is organized reflects my beliefs in making research methods interesting, teaching students how to critique research, and viewing specific research techniques as parts of an integrated research strategy. The text is divided into four sections. The three chapters in the first section, Foundations for Social Research, introduce the why and how of research in general. Chapter 1 shows how research has helped us understand the impact of online social connection. It also introduces some alternative approaches to social research, with a particular emphasis on the contrast between quantitative and qualitative research approaches. Chapter 2 illustrates the basic stages of research with research on the police use of force; I emphasize the role of theory in guiding research and describe the major strategies for research projects. Chapter 3 highlights the importance of research ethics by reviewing the sordid story of the Tuskegee Syphilis Study and its long-term consequences, and I describe the institutional structure developed to maintain adherence to ethical guidelines. The chapter ends by discussing the organization of research proposals.

The next three chapters, in the section on Fundamentals of Social Research, discuss how to evaluate the way researchers design their measures (Chapter 4), draw their samples (Chapter 5), and justify their statements about causal connections (Chapter 6). As you learn about these procedures, you will also read about research on tolerance, controversies about the 2020 U.S. Census, and studies about causes of violence.

In the next section, Basic Social Research Designs, Chapters 7, 8, and 9 present the primary strategies used in quantitative research: collecting data through experiments (Chapter 7) and surveys (Chapter 8) and analyzing data with statistics (Chapter 9). The fascinating research examples in these chapters come from investigations of racial discrimination in the job market, surveys about the pandemic, and the factors associated with voting. Chapters 10 and 11 then introduce the primary strategies used in collecting qualitative data (including participant observation, intensive interviews, and focus groups) and analyzing the results. You will learn in these two chapters about the response to disasters and the strains faced by immigrant families.

The Complex Social Research Designs in Chapters 12 through 16 each involve multiple methodologies. Chapter 12 introduces the mixed methods approaches that combine quantitative and qualitative research techniques to improve understanding of the complexities of the social world, with an example of research on domestic violence during the pandemic. Evaluation research, the focus of Chapter 13, can employ experiments, surveys, and qualitative methods to learn about the need for and the effects of social and other types of programs. This chapter begins with an overview of evaluation research on drug abuse prevention programs. Chapter 14 introduces the techniques and challenges of secondary data analysis and the use of Big Data. The online availability of thousands of data sets from social science studies has helped make secondary data analysis—the use of previously collected data to investigate new research questions—the

method of choice in many investigations. What is termed *Big Data* is generated when you use social media, find your location with GPS, search the web, purchase goods online, or in other ways create electronic records that can be analyzed by researchers to identify social patterns. Examples in this chapter include research on anti-Asian sentiment expressed on Twitter, time use during the pandemic, and social patterns on Facebook.

Chapter 15 focuses on historical and comparative methodologies, which can use data obtained with one or more of the primary methods to study processes at regional and societal levels over time and between units; examples include research on the maintenance of democratic governance and the appeal of authoritarianism. This chapter also introduces demographic methods and content analysis, both of which can be used to good effect in historical and comparative research. Chapter 16 finishes up with an overview of the process of and techniques for reporting research results—with special attention to visual representations of data patterns. A study of international variation in vaccine intentions illustrates the method of meta-analysis, which combines statistically the results of multiple research studies. I also present guidelines for writing about research and discuss and review research on the problem of plagiarism.

DISTINCTIVE FEATURES OF THE TENTH EDITION

The tenth edition of *Investigating the Social World* retains the strengths of previous editions while breaking new ground with the latest developments in research methods, enhanced tools for learning in the text and online, and contemporary, fascinating research findings. The use of research examples reflecting the era of the pandemic, Black Lives Matter, threats to democracy, and online social connection highlights the relevance of social research methods to the progress of the social world. You will find many other innovations in approach, coverage, and organization in the tenth edition.

Updated Coverage of Each Research Method Research methods continue to develop, and new challenges must be overcome as our social world continues to change. Coverage of topics has been updated to reflect changes in the Federal Policy for the Protection of Human Subjects as well as place more emphasis on the following: (1) culturally responsive research ethics, (2) systematic literature reviews, (3) the latest advances in respondent-driven sampling, (4) the challenges of the 2020 U.S. Census, and (5) new research on survey response rates and survey question design. New guidance has been added to the qualitative methods chapters on visual ethnography and on coding qualitative data, while the section on methods for studying the digital social world has been expanded, including an addition of a section on hybrid approaches that combine data from online and offline social interaction. Coverage of the rapidly growing methods for analyzing Big Data has been expanded. In total, insights have been added to the tenth edition from seventeen of SAGE Publications' newest specialty books on particular research methods.

Examples of Social Research as It Occurs in Real-World Settings Every chapter begins with a new example of contemporary research about our tumultuous social world. Fascinating examples of research on vaccination intentions and trust in government, controversy over the 2020 U.S. Census, on police shootings and the effect of body-worn cameras, changes in the homicide rate and in domestic violence during the pandemic, influences on tolerance, effects of climate-related disasters, and cultural change connect research methods to current social issues. Examples also demonstrate the continuing relevance of social theory and the impact of social life on the use of research methods. In total, research examples draw on more than 80 new research

articles in peer-reviewed journals. The *Research in the News* examples have been replaced in each chapter with new vignettes, drawing attention to the role of social research in shaping public discourse and policy.

Adaptation for New Sage Vantage Platform The tenth edition has been redesigned to mesh seamlessly with the SAGE Vantage platform and thus to maximize ease-of-use online as well as in print. Each chapter begins with five to eight learning objectives that in turn correspond to each major section heading in the chapter. The learning objectives adhere to Bloom's Taxonomy and so identify clear, measurable objectives (Anderson and Krathwohl 2001). The learning objectives for each chapter include some requiring lower-order cognitive skills (remembering, understanding, and applying) and some using higher-order cognitive skills (analyzing, evaluating, and creating). Chapter highlights at the end of each chapter are organized in relation to the chapter learning objectives. End-of-chapter learning aids also include Key Terms, Discussion Questions, and SPSS Exercises.

Updated Links and Exercises Exercises have been updated, and those involving IBM SPSS (Statistical Package for the Social Sciences) Statistics* software have been revised for use with the special 2020 General Social Survey (GSS) data set (a version of this data set is available on the book's study site). Chapter 9's introduction to statistical methods can also be used as a tutorial in how to analyze data from the 2020 GSS with SPSS. Versons of the same exercises are available on the study site in the formats required by two other popular statistical packages: R and Stata.

Comprehensive Study Site The many effective study aids included in the previous editions have been updated, as needed. Time you spend on the book's study site will be rewarded with a deeper understanding of research methods and better preparation for exams and papers.

It is a privilege to be able to share with so many students the results of excellent social science investigations of the social world. If *Investigating the Social World* communicates the excitement of social research and the importance of evaluating carefully the methods we use in that research, then I have succeeded in representing fairly what social scientists do. If this book conveys accurately the latest developments in research methods, it demonstrates that social scientists are themselves committed to evaluating and improving their own methods of investigation. I think it is fair to say that we practice what we preach.

Now you're the judge. I hope that you enjoy learning how to investigate the social world and perhaps do some investigating along the way. And I hope you find that the knowledge and (dare I say it?) enthusiasm you develop for social research in this course will serve you well throughout your education, in your career, and in your community.

A NOTE ABOUT USING SPSS*

To carry out the SPSS exercises at the end of each chapter and in Appendix C (on the book's study site), you must already have SPSS on your computer or online (through your university). The exercises use a subset of the 2020 GSS data set (included on the study site). This data set includes variables on topics such as work, family, gender roles, government institutions, race relations, and politics. Appendix C will get you up and running with IBM SPSS, and you can then spend as much time as you like exploring characteristics and attitudes of Americans. If you can use the complete version of SPSS (perhaps in a university computer lab), just download the GSS2020 file and save it on your computer, then start SPSS on your PC, open the GSS2020 file, and

begin with the first SPSS exercise in Chapter 1. If you need to purchase SPSS yourself, the IBM SPSS Statistics Base GradPack 28 is available for Windows and Mac for a student price under $40 (6-month rental). Alternatively, you can use an online analysis program at the University of California, Berkeley, website (http://sda.berkeley.edu/archive.htm) or at the National Opinion Research Center site (https://gss.norc.chicago) to complete many of the SPSS exercises in the text (the ones that do not involve variables created specifically for ISW10). The same exercises are available on the book's study site for students using the statistical packages Stata and R.

ANCILLARIES

This text includes an array of instructor teaching materials designed to save you time and to help you keep students engaged. It also includes access to select student learning resources. To learn more, visit sagepub.com or contact your SAGE representative at sagepub.com/findmyrep.

ACKNOWLEDGMENTS

My thanks first to Helen Salmon, executive editor at SAGE Publishing, who guided planning for the tenth edition during the depths of the pandemic, and to Darcy Scelsi, senior content development editor, whose dedication and clear vision contributed much to the quality of the final text, as did the work of Chelsea Neve in the early development stages. Book production was managed with great expertise and good cheer by Veronica Stapleton Hooper, while the remarkable Karin Rathert again proved herself to be one of the publishing industry's most conscientious and effective copy editors. I am grateful to work with such talented staff at what has become the world's best publisher in social science.

The quality of *Investigating the Social World* draws on the wisdom and creativity of my coauthors on adaptations for other markets and disciplines, as well as from the pleasure of being part of the support group that we provide each other. My profound gratitude to my SAGE coauthors: Ronet Bachman (University of Delaware), Dan Chambliss (Hamilton College), Joe Check (University of Massachusetts Boston), Ray Engel (University of Pittsburgh), and Paul Nestor (University of Massachusetts Boston). And I continue to be grateful for advice shared at our past biennial meetings with coauthors at the Harvard Faculty Club by Philip Brenner, Charles DiSogra, Karen Hacker, Sunshine Hillygus, Peter Marsden, Catherine Kohler Riessman, and Robert J. Sampson.

University of Massachusetts Boston doctoral student Laura R. Golden provided indispensable assistance for the tenth edition. Laura searched the literature for new examples and identified new *Research in the News* stories, checked websites, updated SPSS exercises, and coauthored Chapter 10. She also wrote the instructor's manual for this edition while teaching a research methods class and therefore investigating her own teaching methods. My thanks for her conscientiousness, expertise, and consistent good cheer.

Reviewers for editions one through ten:

Nasrin Abdolali, Long Island University, C. W. Post
Francis O. Adeola, University of New Orleans
Marina A. Adler, University of Maryland, Baltimore
Emmanuel N. Amadi, Mississippi Valley State University
Lynda Ames, State University of New York, Plattsburgh
Doug Anderson, University of Southern Maine
Matthew Archibald, University of Washington
Robert B. Arundale, University of Alaska, Fairbanks
William Augustine, SUNY University at Albany
Karen Baird, Purchase College, State University of New York
Hee-Je Bak, University of Wisconsin–Madison
Von Bakanic, College of Charleston
Carl Bankston, Tulane University
Diane C. Bates, Sam Houston State University

Diana Bates, The College of New Jersey
Andrew E. Behrendt, University of Pennsylvania
Catherine Berheide, Skidmore College
Marit Berntson, University of Minnesota
Karl Besel, Indiana University at Kokomo
Terry Besser, University of Kentucky
Deborah Bhattacharayya, Wittenbert University
Karen Bradley, University of Central Missouri State
Robyn Brown, DePaul University
Cynthia J. Buckley, The University of Texas at Austin
J. P. Burnham, Cumberland College
Jennifer Bulanda, Miami University
Lisa Callahan, Russell Sage College
Gerald Charbonneau, Madonna University
Hugh G. Clark, Texas Woman's University
Mark E. Comadena, Illinois State University
John Constantelos, Grand Valley State University
Sandy Cook-Fon, University of Nebraska at Kearny
Mary T. Corrigan, Binghamton University
Herbert L. Costner, formerly of University of Washington
Jerry Daday, Western Kentucky University
Kelly Damphousse, Sam Houston State University
Ray Darville, Stephen F. Austin State University
James David Ballard, California State University, Northridge
Marvin Dawkins, University of Miami
Robert A. Dentler, University of Massachusetts Boston (Chapter 10)
Jack Dison, Arkansas State University
Qingwen Dong, University of the Pacific
Christopher Donoghue, William Paterson University
Patricia Drentea, University of Alabama at Birmingham
John Eck, University of Cincinnati
Carol Erbes, Old Dominion University
Kristin Espinosa, University of Wisconsin–Milwaukee
Jana Everett, University of Colorado, Denver
Kimberly Faust, Fitchburg State College
Kenneth E. Fernandez, University of Nevada, Las Vegas
Kenneth Fidel, DePaul University
Virginia S. Fink, University of Colorado, Colorado Springs
David H. Folz, University of Tennessee
Gregory Fulkerson, SUNY Oneonta
Sandra K. Gill, Gettysburg College
Lisa AnnGittner, Tennessee State University
Gary Goreham, North Dakota State University
George Guay, Bridgewater State University
Isaac Heacock, Indiana University, Bloomington
Jay Hertzog, Valdosta State University
Jane Hood, University of New Mexico

Lin Huff-Corzine, University of Central Florida
Gary Hytrek, University of California, Los Angeles
Christine Johnson, Oklahoma State University
Joseph Jones, Taylor University
Barbara Keating, Mankato State University
Sean Keenan, Utah State University
Debra S. Kelley, Longwood College
Kurt Kent, University of Florida
Amy Kroska, University of Oklahoma
Manfred Kuechler, Hunter College (CUNY)
Edward Lascher, California State University, Sacramento
Bebe Lavin, Kent State University
Jan Leighley, Texas A&M University
Derek Lester, Texas A&M University
Quan Li, University of Central Florida
Joel Lieberman, University of Nevada, Las Vegas
Carolyn Liebler, University of Washington
Thomas Linneman, College of William & Mary
Andrew London, Kent State University
Scott Long, Indiana University
Stephanie Luce, University of Wisconsin–Madison
Roseanne Macias, California State University at Dominguez Hills
Randall MacIntosh, California State University, Sacramento
Gina Mann-Delbert, Suffolk University
Tajuana D. Massie, South Carolina State University
Peter J. May, University of Washington
Ronald J. McAllister, Elizabethtown College
Steve McDonald, North Carolina State University
Michael McQuestion, University of Wisconsin–Madison
Tricia Mein, University of California at Santa Barbara
Jeanne Mekolichick, Radford University
Carol D. Miller, University of Wisconsin–La Crosse
Elizabeth Monk-Turner, Old Dominion University
Kelly Moore, Barnard College, Columbia University
Bruce Mork, University of Minnesota
Elizabeth Morrissey, Frostburg State University
Chandra Muller, University of Texas
Kevin Mulvey, George Washington University
Kristen Myers, Northern Illinois University
Jennifer R. Myhre, University of California, Davis
Michael R. Norris, University of Texas, El Paso
G. Nanjundappa, California State University, Fullerton
Dan Olson, Indiana University South Bend
Zeynep Özgen, Arizona State University
Jennifer Parker-Talwar, Pennsylvania State University at Lehigh Valley
Nicholas Parsons, Washington State University
Norah Peters-Davis, Beaver College

Jeffrey Prager, University of California, Los Angeles
Ronald Ramke, High Point University
Adinah Raskas, University of Missouri
Liesl Riddle, University of Texas, Austin
Karen Robinson, California State University San Bernardino
Akos Rona-Tas, University of California, San Diego
Janet Ruane, Montclair State University
Josephine A. Ruggiero, Providence College
Jennifer Samson, Arkansas Tech University
David Sanders, Angelo State University
Michael J. Scavio, University of California at Irvine
Mary Ann Schwartz, Northeastern Illinois University
Mildred A. Schwartz, University of Illinois, Chicago (Chapter 11)
Valerie Schwebach, Rice University
Therese Seibert, Keene State College
Shaihid M. Shahidullah, Virginia State University
Tabitha Sharp, Texas Woman's University
Mark A. Shibley, Southern Oregon University
Gi-Wook Shin, University of California, Los Angeles
Pamela J. Shoemaker, Syracuse University
Paul C. Smith, Alverno College
Herbert L. Smith, University of Pennsylvania
Glenna Spitze, University at Albany, State University of New York
Beverly L. Stiles, Midwestern State University
Howard Stine, University of Washington
Judith Stull, Temple University
Brian J. Stults, University of Florida
William J. Swart, The University of Kansas
John Talmage, University of North Florida
Bill Tillinghas, San Jose State University
Jimmy Kazaara Tindigarukayo, University of the West Indies
Carolina Tolbert, Kent State University
Pam Tontodonato, Kent State University
Robbyn Wacker, University of Northern Colorado
Tim Wadsworth, University of Washington
Peggy Walsh, Keene University
Guang-zhen Wang, Russell Sage College
Daniel S. Ward, Rice University
John R. Warren, University of Washington
Greg Weaver, Auburn University
Charles Webb, Freed-Hardeman University
Shernaaz M. Webster, University of Nevada, Reno
Adam Weinberg, Colgate University
Greg Weiss, Roanoke College
DeeAnn Wenk, University of Oklahoma
Karin Wilkins, University of Texas, Austin
Ken Wilson, East Carolina University

Susan Wurtzburg, University of Hawai'i
Keith Yanner, Central College
Kevin A. Yoder, University of North Texas

I am also grateful for Kathy Crittenden's support on the first three editions, for the contributions of Herbert L. Costner and Richard Campbell to the first edition, and to Steve Rutter, whose vision and enthusiasm launched the whole project on a successful journey. Special thanks to Barbara Costello, University of Rhode Island; Nancy B. Miller, University of Akron; and Gi-Wook Shin, University of California, Los Angeles, for their contributions to the third edition.

Philip Brenner provided helpful feedback on Chapter 8 for the ninth edition, as did Reef Youngreen and Phil Kretsedemas for the seventh edition on, respectively, Chapters 3 and 4. Several former faculty, staff, and graduate students at the University of Massachusetts Boston made important contributions to earlier editions: Heather Albertson, Ra'eda Al-Zubi, Bob Dentler and students in his 1993–1994 graduate research methods class, as well as Anne Foxx, Whitney Gecker, Chris Gillespie, Tracey Newman, Megan Reynolds, Kathryn Stoeckert, Tatiana Williams-Rodriguez, and Jeffrey Xavier. Heather Johnson at Northeastern University also contributed to an earlier edition. I continue to be indebted to the many students I have had the opportunity to teach and mentor, at both the undergraduate and graduate levels. In many respects, this book could not have been so successful without the ongoing classroom experiences we have shared. I also share a profound debt to the many social scientists, medical scientists, and service professionals with whom I have collaborated in social science research projects.

No scholarly book project can succeed without good library resources, and for these I continue to incur a profound debt to the Harvard University library staff and their extraordinary collection. I also have benefited from the resources maintained by the excellent librarians at the University of Massachusetts Boston.

Again, most important, I thank my wife, Elizabeth, for her love and support, and our daughter, Julia, and niece, Enengy, for the joy they bring to our lives and their own contributions in the social world.

—Russell K. Schutt

SECTION I

FOUNDATIONS FOR SOCIAL RESEARCH

Chapter 1	Science, Society, and Social Research	2
Chapter 2	The Process and Problems of Social Research	31
Chapter 3	Research Ethics and Research Proposals	75

©iStockphoto.com/scyther5

1 SCIENCE, SOCIETY, AND SOCIAL RESEARCH

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

1.1 Describe three common types of motivations for social research

1.2 Illustrate each of the four common errors involved in learning about the social world using a hypothetical example

1.3 List components of the scientific approach

1.4 Compare the four types of social research

1.5 Assess the strengths and limitations of social research

1.6 Contrast three pairs of alternatives in social research: positivism and postpositivism; quantitative methods and qualitative methods; and basic research and applied research

How do social connections through the Internet compare to those that are face-to-face? Are they as meaningful? As satisfying? As helpful? Have smartphones, the World Wide Web, and social media changed our social world for the better? Questions like this motivated a lot of social science research after the launch of the World Wide Web in 1991. Then in Spring 2020, COVID-19 swept across the globe and most normal social activities, including social science research, came to an abrupt halt. Unparalleled levels of illness and death ruptured social connections and redrew social divisions. We have all been changed.

We have also participated in an unanticipated experiment about online social connection. How did you manage? Did you live for weeks or months without any meaningful in-person contact with others, at least outside your immediate family or roommates? Did your Internet connections work well for you and those you wanted to be in touch with? Did your time online change your relations with others, whether you saw them in person or only online? How did your family members, friends, and neighbors manage?

That's where social researchers begin, with questions about the social world and a desire to answer them. What makes social research different from the ordinary process of thinking about our experiences is a focus on broader questions that involve people outside our immediate experience, questions about why things happen that we may not otherwise consider, and the use of systematic research methods to answer those questions. It's just as true now, as we reflect on the impact of the pandemic. Each of us has had many relevant personal experiences, ranging from feelings about social distancing and losing jobs or income to perceptions of others' behaviors and images of others' suffering. How widespread are our own feelings and how accurate are our perceptions? What has changed because of the pandemic, and what else influenced people during the pandemic? Social research methods help us to answer these bigger questions, challenge our preexisting assumptions, and develop more confidence in our answers. Social scientists think of the big picture, try not to neglect the details, strive to figure out how the pieces go together, and keep a record at each step.

And that's just for starters. In this chapter, I hope to convince you that the use of social science research methods to investigate questions about the social world results in knowledge that can be more important, more trustworthy, and more useful than only

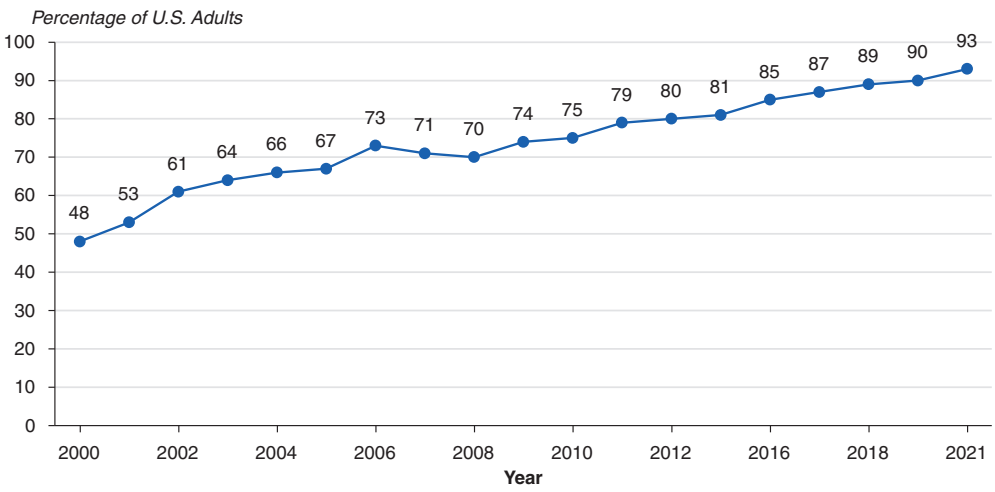
relying on personal opinions, individual experiences, or someone's Twitter posts. You will also learn about the challenges that researchers confront and the standards to which they must adhere. By the chapter's end, you should know what is "scientific" in social science and appreciate how the methods of science can help us understand the problems of society.

LEARNING ABOUT THE SOCIAL WORLD

We can begin to get a sense of how sociologists and other social scientists investigate the social world by reviewing some questions that social researchers have asked about the Internet and social ties and the ways they have answered those questions before and during the pandemic.

1. *How connected are people to the Internet?* The Pew Research Center's surveys have found that Internet use in the United States has risen rapidly, from 52% of U.S. adults in 2000 to 93% in 2021 (Perrin and Atske 2021) (see Figure 1.1). By the end of March 2021, 65.6% of the total world population of 7,875,765,587 was connected in some way to the Internet—an increase of 1,332% since 2000. Across continents, the percentage connected ranged from highs of 93.9% in North America and 88.2% in Europe to 74.9% in the Middle East, 69.9% in Oceania and Australia, 63.8% in Asia and 43.2% in Africa (Internet World Stats 2021). These connections increased rapidly during the pandemic, with Internet usage rising by 40% or more and videoconferencing increasing ten times (De, Pandey, and Pal 2020). How did your own Internet usage change?

FIGURE 1.1 ■ Percentage of U.S. Adults Who Use the Internet, 2000–2021



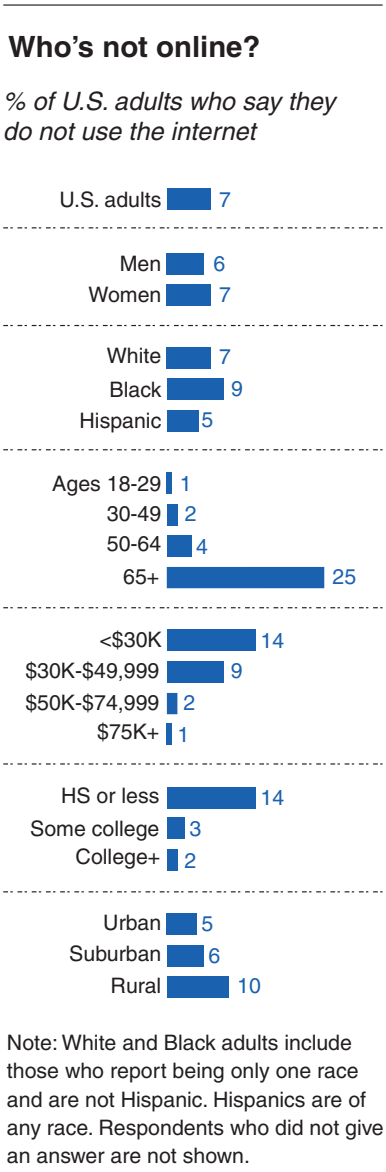
Source: "7% of Americans Don't Use the Internet. Who Are They?" Pew Research Center, Washington, DC (April 2, 2021) https://www.pewresearch.org/fact-tank/2021/04/02/7-of-americans-dont-use-the-Internet-who-are-they/ and "Americans' Internet Access: 2000–2015." Pew Research Center, Washington, DC. (June 26, 2015) http://www.pewInternet.org/2015/06/26/americans-Internet-access-2000-2015/.

2. *How does Internet use vary across social groups?* Pew's surveys have also revealed differences in Internet use between social groups. The percentage of U.S. adults who were not online in 2021 (7% overall) was similar between men and women, differed

only a bit by race, but varied markedly by age—from a low of 1% of those aged 18–29 to a high of 25% among those 65 or older. Internet use also varied by income, education, and location, with those in the lowest income and educational groups least likely to be online (Perrin and Atske 2021) (see Figure 1.2). How much variation have you observed across people with these different characteristics?

3. *How did the COVID-19 pandemic influence Internet use?* A large majority of U.S. adults said in April 2020 that the Internet was "essential" (53%) or at least "important"

FIGURE 1.2 ■ Percentage of Individuals Not Using Internet, by Personal Characteristics


Who's not online?
% of U.S. adults who say they do not use the internet
U.S. adults 7
Men 6
Women 7
White 7
Black 9
Hispanic 5
Ages 18-29 1
30-49 2
50-64 4
65+ 25
<$30K 14
$30K-$49,999 9
$50K-$74,999 2
$75K+ 1
HS or less 14
Some college 3
College+ 2
Urban 5
Suburban 6
Rural 10
Note: White and Black adults include those who report being only one race and are not Hispanic. Hispanics are of any race. Respondents who did not give an answer are not shown.

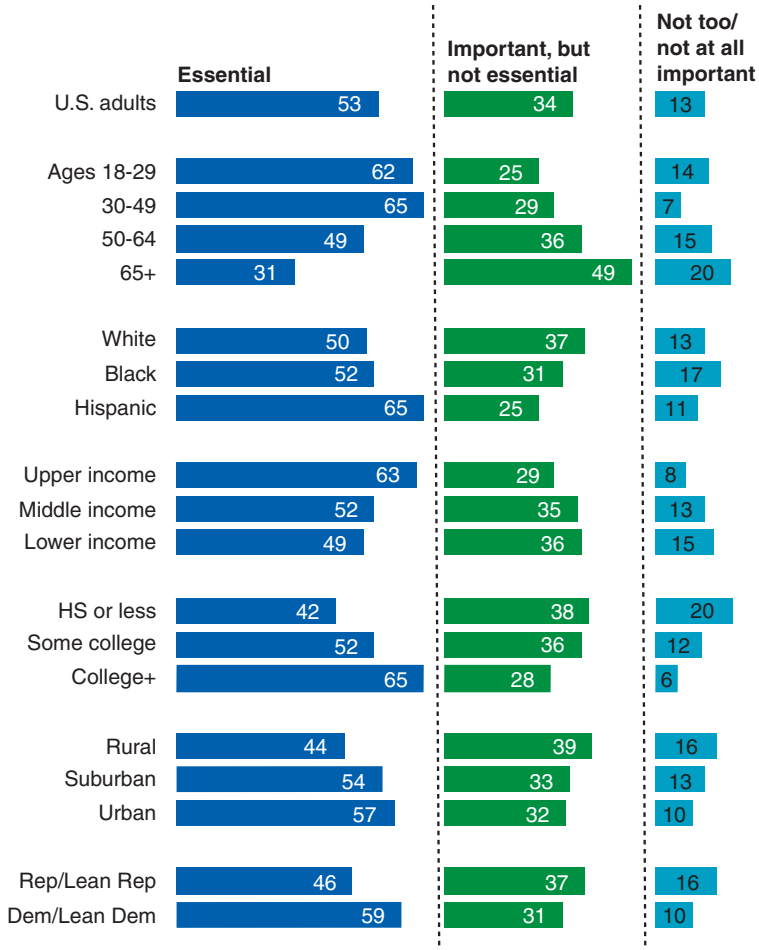
Source: "7% of Americans Don't Use the Internet. Who Are They?" Pew Research Center, Washington, DC (April 2, 2021) https://www.pewresearch.org/fact-tank/2021/04/02/7-of-americans-dont-use-the-internet-who-are-they/ and "Americans' Internet Access: 2000–2015." Pew Research Center, Washington, DC. (June 26, 2015) http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/.

(34%) to them during the pandemic. Adults under 50, Hispanics, those with high incomes and with a college education rated the Internet as more important to them during the pandemic compared to those older, not Hispanic, and having less education and income. Those in urban areas and those who identified more as Democratic also rated the Internet as more important than their counterparts (see Figure 1.3). Was the Internet essential for you during the pandemic?

FIGURE 1.3 ■ Importance of Internet During Pandemic by Social Characteristics

Hispanic adults, college graduates and adults under 50 are especially likely to say the internet has been essential during the coronavirus outbreak

% of U.S. adults who say the internet has been ___for them personally during the coronavirus outbreak


Essential
Important, but not essential
Not too/ not at all important
U.S. adults 53 34 13
Ages 18-29 62 25 14
30-49 65 29 7
50-64 49 36 15
65+ 31 49 20
White 50 37 13
Black 52 31 17
Hispanic 65 25 11
Upper income 63 29 8
Middle income 52 35 13
Lower income 49 36 15
HS or less 42 38 20
Some college 52 36 12
College+ 65 28 6
Rural 44 39 16
Suburban 54 33 13
Urban 57 32 10
Rep/Lean Rep 46 37 16
Dem/Lean Dem 59 31 10

Note: White and black adults include only those who are non-Hispanic. Hispanics are of any race. Family income tiers are based on adjusted 2018 earnings. Those who did not give an answer are not shown.

Source: Vogels, Emily, Andrew Perrin, Lee Rainie, and Monica Anderson. April 2020. "53% of Americans Say the Internet Has Been Essential During the COVID-19 Outbreak." Washington, DC: Pew Research Center.

4. *How well did online social connections substitute for in-person social interaction?* Chad E. Kee (2021) sought to answer this question by adding a research component to his graduate class in higher education leadership at Morgan State University in Baltimore. Systematic analysis of student reflection papers during the Spring 2020 semester identified five main themes, including "experiencing loss of power and control"—reflected in statements like, "In this new reality, work-life balance has become work-life integration and it's definitely harder for me to compartmentalize"—and "managing feelings of anxiety and fear"—as one student described it, "an overwhelming concern about how we are going to 'get back to normal,' and panic that we may not be able to." Sound familiar? The Pew Research Center (2020) found that younger adults and those with more education were much more likely to have had a social gathering online with friends or family during the pandemic than those who were 50 or above or had less than a college degree. Did you get together with friends or family online? How did that experience compare to in-person get-togethers before the pandemic?

Did your personal experiences lead you to expect different answers to these questions? You have just learned that younger people use the Internet more than do older people. Does this variability lead you to be cautious about using your own experience as a basis for estimating the behavior of others (#2)? Have you heard others say the Internet became essential for them during the pandemic? Is it safe to draw general conclusions from this anecdotal evidence (#3)? Have you observed differences in how satisfying electronic forms of communication have been to different people (#4)?

We cannot avoid asking questions about our complex social world or trying to make sense of our position in it. The more that you begin to "think like a social scientist," the more such questions will come to mind—and that's a good thing! But in our everyday reasoning about the social world our own prior experiences and orientations can have a major influence on what we perceive and how we interpret these perceptions. As a result, one person may see someone posting a message on Facebook as being typical of what's wrong with modern society, but another person may see the same individual as helping people "get connected" with others. We need to move beyond first impressions and gut reactions to more systematic methods of investigation.

Motives for Social Research

Similar to you, social scientists have friends and family, observe other persons' social ties, and try to make sense of what they experience and observe. For most, that's the end of it. But for some social scientists, the quality and impact of social ties has become a major research focus. What motivates selection of this or any other particular research focus? Usually, it's one or more of the following reasons:

- *Policy motivations.* Many government agencies, elected officials, and private organizations seek better descriptions of social ties in the modern world so they can identify unmet strains in communities, deficits in organizations, or marketing opportunities. Public health officials may need information for planning vaccination campaigns in different communities. Law enforcement agencies may seek to track the connections between criminal gangs and the effect of social cohesion on the crime rate. Military leaders may seek to strengthen unit cohesion. These policy guidance and program management needs can stimulate numerous

research projects. As Kathleen Cooper and Nancy Victory (2002) said in their foreword to a U.S. Department of Commerce report on the Census Bureau's survey of Internet use,

> This information will be useful to a wide variety of policymakers and service providers. . . help all of us determine how we can reach Americans more effectively and take maximum advantage of the opportunities available through new information technologies. (p. iii)

- Academic *motivations.* Questions about changing social relations have stimulated much academic social science. More than one hundred years ago, Émile Durkheim (1951/1897) linked social processes stemming from urbanization and industrialization to a higher rate of suicide. More than fifty years ago, David Reisman (1969/1950) considered whether the growing role of the mass media, among other changes, was leading Americans to become a "lonely crowd." Similar to this earlier research, contemporary investigations of the effect of computers and the Internet are often motivated by a desire to understand influences on the strength and meaning of social bonds. Did online social connections during the pandemic perform the same functions as face-to-face social relationships (Schutt et al.)? The desire to understand better how the social world works is motivation enough for many social scientists (Hampton and Wellman 2001):

  > It is time to move from speculation to evidence. . . . The growth of computer-mediated communication (CMC) introduces a new means of social contact with the potential to affect many aspects of personal communities. . . . This article examines . . . how this technology affected contact and support. (pp. 477, 479)

- *Personal motivations.* Some social scientists who conduct research on social ties feel that by doing so they can help improve the quality of communities, the effectiveness of organizations, or the physical and mental health of many social groups. Social scientists may become interested in social ties as a result of exposure to problems in the social world or by watching the challenges their children face learning remotely or for many other reasons, including finding themselves without many friends after a career move. Can you imagine a college student, in later years, developing an interest in research on poverty in other countries as a result of a study abroad experience?

AVOIDING ERRORS IN REASONING ABOUT THE SOCIAL WORLD

How can we avoid errors rooted in the particularities of our own backgrounds and improve our reasoning about the social world? The pandemic exposed deep differences between people in perceptions about COVID-19's spread and in reasoning about its social implications. Charges of "fake news" and challenges of "expert authority" served as a constant reminder of how important it is to avoid errors and improve our reasoning. Let's begin by identifying the different processes involved in learning about the social world and the types of errors that can result as we reason about the social world.

When we learn about the social world, we engage in one or more of four processes: (1) *observing* through our five senses (seeing, hearing, feeling, tasting, or smelling); (2) *generalizing* from what we have observed to other times, places, or people; (3) *reasoning* about the connections between

different things we have observed; and (4) *reevaluating* our understanding of the social world on the basis of these processes. It is easy to make mistakes with each of these processes. It's also important to realize that we learn about the social world while we are participating in it. We often observe the social world indirectly, through images and actions we see in the media—ranging from Tweets to TV—and our generalizing, reasoning, and reevaluating always reflect the thoughts and feelings of others—ranging from those who have taught us in the past to those who seek to convince us in the present.

My favorite example of the errors in reasoning that occur in the nonscientific, unreflective discourse about the social world that we hear on a daily basis comes from a letter to famous advice columnist Ann Landers. The letter was written by someone who had just moved with her two cats from the city to a house in the country. In the city, she had not let her cats outside and felt guilty about confining them. When they arrived in the country, she threw her back door open. Her two cats cautiously went to the door and looked outside for a while, then returned to the living room and lay down. Her conclusion was that people shouldn't feel guilty about keeping their cats indoors—even when they have the chance, cats don't really want to play outside. Do you see this person's errors in her approach to:

- *Observing?* She observed the cats at the outside door only once.
- *Generalizing?* She observed only two cats, both of which previously were confined indoors.
- *Reasoning?* She assumed others feel guilty about keeping their cats indoors and that cats are motivated by feelings about opportunities to play.
- *Reevaluating?* She was quick to conclude that she had no need to change her approach to the cats.

You don't have to be a scientist or use sophisticated research techniques to avoid these four errors in reasoning. If you recognize these errors for what they are and make a conscious effort to avoid them, you can improve your own reasoning about the social world. In the process, you will also be implementing the admonishments of your parents (or minister, teacher, or any other adviser) to avoid stereotyping people, to avoid jumping to conclusions, and to look at the big picture. These are the same errors the methods of social science are designed to help us avoid. We need to need to follow the advice to think for ourselves, but we also have to be open to critical examination of our perceptions and thoughts.

Observing

One common mistake in learning about the social world is **selective observation**—choosing to only look at or listen to things that are in line with our preferences or beliefs. The disagreements during the early months of the pandemic over hydroxychloroquine as a preventive treatment provided striking examples of the problem.

In a May 2020 cabinet meeting, former president Trump was quoted as saying,

> Hydroxychloroquine is used by thousands and thousands of front-line workers, so that hopefully they don't catch this horrible disease or whatever you want to call it.

But the American Medical Association reported that it "does not know of any tracking or surveys examining personal use of hydroxychloroquine among health workers." The American

Nurses Association said it "has not received reports from nurses or other front-line health care workers utilizing hydroxychloroquine as a preventative treatment for COVID-19." However, *The New York Times* (Qiu May 2020) story revealing these events also reported earlier in the year that some doctors had been taking and hoarding hydroxychloroquine before the Food and Drug Administration issued a warning about it and studies questioned its value. There were also thousands of health care and clinical workers participating in clinical trials to test the value of hydroxychloroquine.

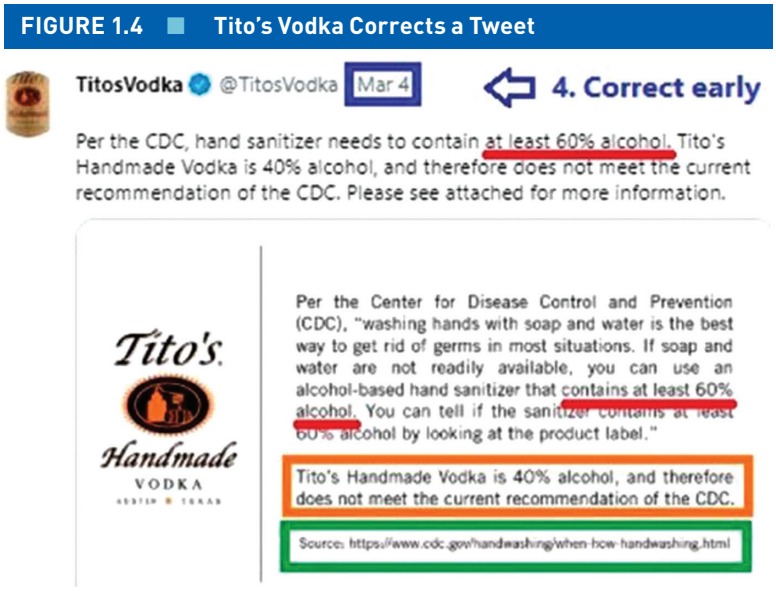
So the former president's remarks seemed to represent observations selected to support his position rather than an attempt to observe all of the available evidence. The lesson is that if we acknowledge only the instances that confirm our predispositions, we are victims of our own selective observation.

Our observations can also simply be inaccurate. If, after a quick glance inside the local pharmacy, you think there are 14 persons waiting in the vaccine line, when there are actually 17, you have made an **inaccurate observation**. If you hear a speaker say that "for the oppressed, the flogging never really stops," when what she said was, "For the obsessed, the blogging never really stops" (Hafner 2004), you have made an inaccurate observation.

Tito's Vodka quickly corrected a claim made in someone's Tweet based on an inaccurate observation about the use of their product as a hand sanitizer (Figure 1.4). Partisan politicking has made such fact-checking a very big and very necessary business.

Such errors occur often in casual conversation and in everyday observation of the world around us. In fact, our perceptions do not provide a direct window onto the world around us, for what we think we have sensed is not necessarily what we have seen (or heard, smelled, felt, or tasted). Even when our senses are functioning fully, our minds have to interpret what we have sensed (Humphrey 1992). The optical illusion in Figure 1.5, which can be viewed as either two faces or a vase, should help you realize that perceptions involve interpretations. Different observers may perceive the same situation differently because they interpret it differently.

FIGURE 1.4 ■ Tito's Vodka Corrects a Tweet


TitosVodka @TitosVodka Mar 4
4. Correct early
Per the CDC, hand sanitizer needs to contain at least 60% alcohol. Tito's Handmade Vodka is 40% alcohol, and therefore does not meet the current recommendation of the CDC. Please see attached for more information.
Tito's Handmade VODKA
Per the Center for Disease Control and Prevention (CDC), "washing hands with soap and water is the best way to get rid of germs in most situations. If soap and water are not readily available, you can use an alcohol-based hand sanitizer that contains at least 60% alcohol. You can tell if the sanitizer contains at least 60% alcohol by looking at the product label."
Tito's Handmade Vodka is 40% alcohol, and therefore does not meet the current recommendation of the CDC.
Source: https://www.cdc.gov/handwashing/when-how-handwashing.html

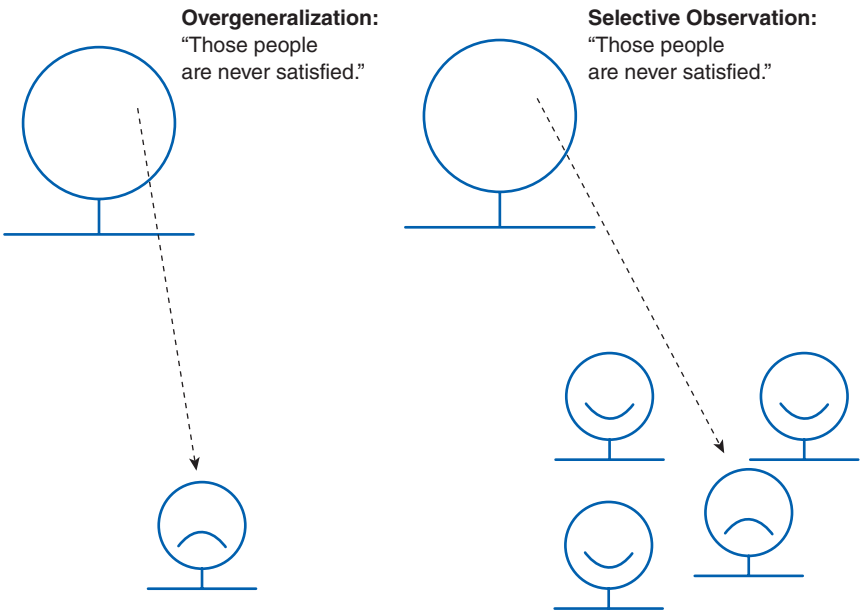
Source: ©Twitter.com

FIGURE 1.5 ■ An Optical Illusion

Generalizing

Overgeneralization occurs when we conclude that what we have observed or what we know to be true for some cases is true for many more or all cases (see Figure 1.6). We are always drawing conclusions about people and social processes from our own interactions with them and perceptions of them or based on information about a limited number of examples. Overgeneralization was part of the problem after the CDC identified 17 cases (ultimately, a total of 28) of potentially lethal blood clots among those who had received the Johnson & Johnson COVID-19 vaccine (Miller and Reuters 2021, "Johnson & Johnson vaccine linked to 28 cases of blood clots, CDC

FIGURE 1.6 ■ The Difference Between Selective Observation and Overgeneralization



reports" [nbcnews.com]). Acceptance of the vaccine plummeted after this pause, even though more than 8 million people had received the vaccine without complications, which results in a tiny risk of 1.9 cases per million (Krieger May 2021, "The fall from grace for Johnson & Johnson's COVID-19 vaccine" [mercurynews.com]; MIT Medical May 2021, "J&J is a-ok". [| MIT Medical]). The fear that the harm might generalize to many others could not be easily dispelled.

Thanks to the Internet and social media, we can easily find many other examples of overgeneralization. Here's one posted by a frequent blogger who was called for jury duty (http://busblog.tonypierce.com/2005/06/yesterday-ihad-to-go-to-jury-duty-to.html, posted on June 17, 2005):

> yesterday i had to go to jury duty to perform my civil duty. *unlike most people* i enjoy jury duty because i find the whole legal process fascinating, especially when its unfolding right in front of you and you get to help decide yay or nay.

Do you know what the majority of people think about jury duty? According to a Harris Poll, 75% of Americans consider jury service to be a privilege (Grey 2005), so the blogger's generalization about "most people" is not correct. Do you ever find yourself making a quick overgeneralization like that?

Reasoning

When we prematurely jump to conclusions or argue on the basis of invalid assumptions, we are using **illogical reasoning**. Verbal and physical assaults on Asian Americans since the pandemic began provide many examples of such illogical reasoning (as well as of disregard for others).

> A 30-year-old videographer in Syracuse, N.Y., said he was still shaken from a trip to the grocery store last week, when the man ahead of him in the checkout line shouted at him, "It's you people who brought the disease." (Tavernise and Oppel 2021. Spit On, Yelled At, Attacked: Chinese-Americans Fear for Their Safety (nytimes.com).

Is it logical to think that a woman walking to the gym or a person who directs a hospital emergency room or a magazine writer "brought the disease" because they have Asian backgrounds? Do you know the so-called "Spanish flu" that caused the 1918 pandemic did not originate in Spain at all, but most likely in a town in Kansas (Barry 2005)? Would it have been logical for people in other countries to start attacking Americans for having "brought the flu" as it spread around the world?

It is not always so easy to spot illogical reasoning. As you learned above, by 2021, only 7% of American households reported not using the Internet. Would it be reasonable to propose they don't participate in the "information revolution" simply because they think it is a waste of time? In fact, many low-income households lack the financial resources to buy a computer or maintain an online account and so they use the Internet much less frequently—as you saw in Figure 1.2. Conversely, an unquestioned assumption that everyone wants to connect to the Internet may overlook some important considerations; for example, 17% of nonusers of the Internet in 2002 said the Internet has made the world a worse place, so they may not use it because they don't like what they believe to be its effects (UCLA Center for Communication Policy 2003:78). Logic that seems impeccable to one person can seem twisted to another.

Reevaluating

Resistance to change, the reluctance to reevaluate our ideas in light of new information, may occur for several reasons:

- *Ego-based commitments.* We all learn to greet with some skepticism the claims by leaders of companies, schools, agencies, and so on that people in their organization are happy, that revenues are growing, and that services are being delivered in the best possible way. We know how tempting it is to make statements about the social world that conform to our own needs rather than to the observable facts. It can also be difficult to admit we were wrong once we have staked out a position on an issue. Barry Wellman (Boase et al. 2006:1) recounts a call from a reporter after the death of four "cyber addicts." The reporter was already committed to the explanation that computer use had caused the four deaths; now, he just wanted an appropriate quote from a computer-use expert, such as Wellman. But the interview didn't last long:

  > The reporter lost interest when Wellman pointed out other causes might be involved, that "addicts" were a low percentage of users, and no one worries about "neighboring addicts" who chat daily in their front yards. (Boase et al. 2006:1)

- *Excessive devotion to tradition.* Some degree of devotion to tradition is necessary for the predictable functioning of society. Social life can be richer and more meaningful if it is allowed to flow along the paths charted by those who have preceded us. Some skepticism about the potential for online learning once served as a healthy antidote to unrealistic expectations of widespread student enthusiasm (Bray 1999). But too much devotion to tradition can stifle adaptation to changing circumstances. When we distort our observations or alter our reasoning so we can maintain beliefs that "were good enough for my grandfather, so they're good enough for me," we hinder our ability to accept new findings and develop new knowledge. Of course, there was nothing "traditional" about maintaining social ties through e-mail when this first became possible in the late 20th century. Many social commentators assumed the result of increasing communication by e-mail would be fewer social ties maintained through phone calls and personal contact. As a result, it was claimed, the social world would be impoverished. But subsequent research indicated that people who used e-mail more also kept in touch with others more in person and by phone (Benkler 2006:356; Boase et al. 2006).

- *Uncritical agreement or disagreement with authority.* If we do not have the courage to evaluate critically the ideas of those in positions of authority or, conversely, to consider that their statements could ever be right, we will have little basis for complaint if they exercise their authority over us in ways we don't like or if we fail to reap the benefits of decisions that could benefit us. And, if we do not allow new discoveries to challenge our beliefs, our understanding of the social world will remain limited. Was it partly uncritical agreement with authorities "on our side" that led to such disagreements about Russia's role in the 2016 presidential election? A March 2018 poll in the United States found that 82% of Democrats believed Russia spread fake news to help Donald Trump win the 2016 election (agreeing with U.S. intelligence agencies), and 66% believed Russia tampered with the vote tallies (without any evidence). Conversely, just 29% of Republicans agreed Russia spread fake news stories to help Trump (disagreeing with the intelligence agencies), and just 18% believed Russia tampered with the votes (agreeing with the intelligence agencies) (Nyhan 2021).

Now take just a minute to reexamine the issues about social ties, Internet use, and even the 2016 election. Did you grasp at a simple explanation even though reality is far more complex?

Did your own ego and feelings about your similarities to or differences from others influence your beliefs? Did you weigh carefully the opinions of people with access to information you didn't have? Could knowledge of research methods help improve your own understanding of the social world? Do you see some of the challenges social science faces?

SCIENCE AND SOCIAL SCIENCE

Science relies on logical and systematic methods to answer questions, and it does so in a way that allows others to inspect and evaluate its methods. In this way, scientific research develops a body of knowledge that is continually refined, as beliefs are rejected or confirmed on the basis of testing empirical evidence.

"Trust the science" was often used to justify policy decisions during the pandemic, but in such a polarized political period, such "trust" often meant different things to different people. An international survey in the year before the pandemic began found that just over one third of residents of the United States and 19 other developed nations have "a lot" of trust that scientists do what is right for the public; most of the rest have "some" trust (39% in the U.S.) rather than "not too much" or no trust (21% in the U.S.) (Funk, Kennedy, and Johnson 2020).

The value that scientific research has added to the human experience ranges from the clothes we wear to the vehicles we use and the smartphones we have and the vaccines that protect us. So why isn't there more trust in science? One reason is that science also provided the knowledge for nuclear weapons, climate change, and tools for invading privacy. It's not entirely surprising this legacy makes some people suspicious.

Another reason is that scientific knowledge is continually refined and so can change over time. For example, early in the pandemic, experience with other viruses led scientists at the U.S. Centers for Disease Control and Prevention (CDC) to believe that COVID-19 could be spread easily by touching contaminated surfaces. When this proved to be wrong after additional evidence was collected, guidelines were, appropriately, changed. This doesn't mean the initial belief was "not scientific" or the scientific method doesn't work (van der Bles et al. 2020). Science is an ongoing, cumulative process. Would masks help to prevent infection with COVID-19? Early in the pandemic, the body of scientific evidence based on other epidemics or pandemics led many to believe mask-wearing would not help (Greenhalgh et al. 2020). But new evidence on the COVID-19 pandemic rapidly changed scientists' beliefs about that.

RESEARCH IN THE NEWS

Science of Virus Keeps Evolving and Frustrating

"Americans are living with science as it unfolds in real time," *New York Times* health and science reporter Apoorva Mandaville observed. We have all looked eagerly for the latest news on rates of infection with the coronavirus, evidence of the effects of infection, and ways to stay safe. Yet the findings of scientific research have not been simple or consistent, whether they have come from the CDC, the WHO, university-based researchers, or others. Research-based guidelines have changed during the pandemic about contaminated surfaces, about the frequency of viral evolution, and about mask wearing. Experts have disagreed among themselves at times on some of these issues, leaving some Americans feeling "bewildered and bamboozled" and disillusioned with science.

Mandaville points to popular lack of understanding of science as a major source of such disillusion. Science normally progresses gradually to an improved understanding of complex

physical and social processes through multiple investigations by many different scientists who share findings at conferences and criticize and evaluate each other's work prior to publication and public release. This process has been short-circuited due to the need to develop public guidelines as quickly as possible in the midst of a deadly, rapidly spreading viral threat. Mandaville suggests that health officials need to counter the "pseudoscience and lies" spread through social media about the pandemic by correcting misunderstandings about science and clearly communicating current uncertainties when new findings are publicized.

For Further Thought?

1. Have you and your friends become more or less confident in science during the pandemic?
2. How could social science research identify better means for communicating scientific findings?

News source: Mandavilli, Apoorva. 2021. "Science of Virus Keeps Evolving and Frustrating." *The New York Times*, August 23:A1, A17.

The Scientific Approach

The sciences—physical (e.g., physics), natural (e.g., biology), social (e.g., sociology)—share an approach developed over centuries and was formalized in the 17th and 18th centuries by philosophers such as Francis Bacon (1561–1626) and early scientists, including Galileo (1564–1642) and Newton (1643–1727). The central element of this approach (often called the "scientific method" or the "scientific attitude") is investigating phenomena in the world by testing ideas about them with observations of those phenomena—empirical data. Will two stones of different weight fall at the same or different speeds? Test it out—by dropping them (Galileo). Is white light composed of a mixture of different colors? Test it out—with a prism (Newton).

We can summarize this central element of the scientific approach with two words:

- *Empiricism.* Science focuses on phenomena in the real world that can be observed directly or indirectly.
- *Objectivity.* The goal of science is an objective assessment of evidence—freedom from bias due to personal background or pressures. What this means in practice is a commitment to "intersubjectivity": accepting as scientifically trustworthy only evidence that receives support from other scientists.

Most scientists seek to maintain several other standards to increase their ability to conduct research that is empirical and objective (or intersubjective) (Grinnel 1992; Popper 1968; Wallace 1983):

- *Falsifiability.* Scientific ideas can be put to a test and potentially shown to be false.
- *Theoretical.* Science seeks general explanations for phenomena.
- *Community.* Scientific research is conducted in a community of scientists who share and challenge each other's findings and beliefs.
- *Simplicity.* A proposed explanation is preferred over others if—other things being equal—it is simpler.

The basic idea is that scientists test their explanations (falsifiability) in ways that stand up to critique by other scientists (community) and that make sense in relation to other scientific

findings (theoretical) without being more complicated than necessary (simplicity). We will return to the scientific approach in more detail in the next chapter.

Social science relies on scientific methods to investigate individuals, societies, and social processes. It is important to realize that when we apply scientific methods to understanding ourselves, we often engage in activities—asking questions, observing social groups, or counting people—that are similar to things we do in our everyday lives. However, social scientists develop, refine, apply, and report their understanding of the social world more systematically, or "scientifically," than Joanna Q. Public—or some who just tweet their opinions or post a compelling blog entry.

- Social science research methods can reduce the likelihood of overgeneralization by using systematic procedures for selecting individuals or groups to study that are representative of the individuals or groups to which we want to generalize.
- To avoid illogical reasoning, social researchers use explicit criteria for identifying causes and for determining whether these criteria are met in a particular instance.
- Social science methods can reduce the risk of selective or inaccurate observation by requiring that we measure and sample phenomena systematically.
- Because they require that we base our beliefs on evidence that can be examined and critiqued by others, scientific methods lessen the tendency to develop answers about the social world from ego-based commitments, excessive devotion to tradition, or unquestioning respect for authority.

Even as you learn to appreciate the value of social science methods, however, you shouldn't forget that *social* scientists face three specific challenges:

1. The objects of our research are people like us, so biases rooted in our personal experiences and relationships are more likely to influence our conclusions than if we are studying stars in the sky or weeds in the sea.
2. Those we study can evaluate us, even as we study them. As a result, research participants' decisions to "tell us what they think we want to hear" or, alternatively, to refuse to cooperate in our investigations can produce misleading evidence.
3. In physics or chemistry, research subjects (objects and substances) may be treated to extreme conditions and then discarded when they are no longer useful. However, social (and medical) scientists must concern themselves with the way their human subjects are treated in the course of research (much could also be said about research on animals, but this isn't the place for that).

The current period of heightened political polarization has increased these challenges. It is too easy to quickly categorize statements as on "our side" or "their side" and then to judge their accuracy accordingly (Druckman et al. 2021). Even beliefs in the trustworthiness of science are polarized in the United States, with 62% of those who lean ideologically left trusting scientists a lot, compared to just 20% of those who lean ideologically right (Funk et al. 2020). As one result, Democrats were much more likely to see COVID-19 as a major threat to the health of the U.S. public—as did public health scientists—than Republicans (Tyson 2021). Understanding the scientific method is not just an academic matter!

Unfortunately, we cannot assume that others understand the scientific method or much less, that evidence generated in scientific investigations should be valued. Sage advice about this in relation to another politically polarized issue—police shootings—was offered by Dr. Lawrence W. Sherman, Wolfson Professor of Criminology Emeritus at the University of Cambridge Institute of Criminology and former president of the American Academy of Political and Social Science as well as former distinguished university professor at the University of Maryland.

> It is, after all, irrational to think that better evidence should win out, just because it is scientifically better than the alternative. Like police legitimacy, better evidence must sell itself in a public dialogue, every day, even with setbacks.

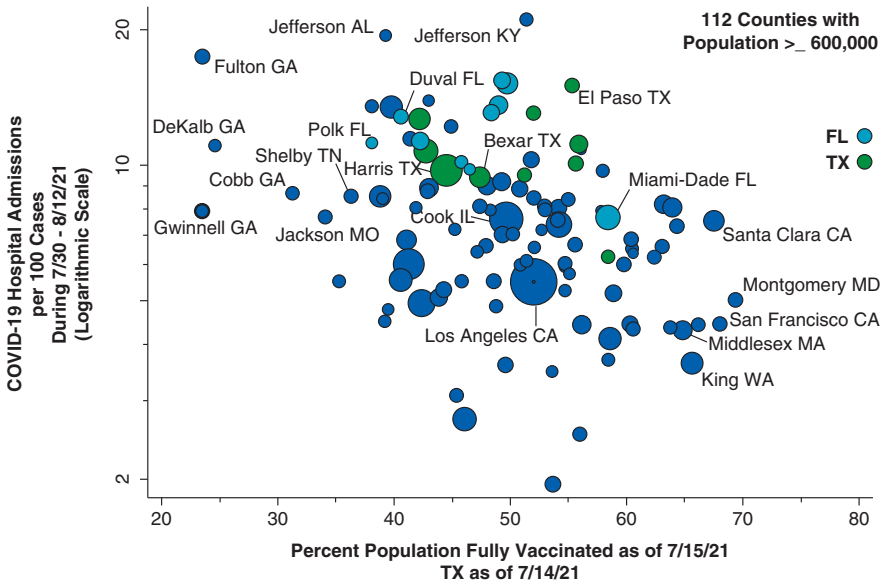
Pseudoscience or Science

We must also be on guard against our natural tendency to be impressed with knowledge that is justified with what sounds like scientific evidence but which has not really been tested. **Pseudoscience** claims are not always easy to identify, and many people believe them (Shermer 1997:33).

Are you surprised that more than half of Americans believe in astrology, with all its charts and numbers and references to stars and planets, even though astrological predictions have been tested and found baseless (Shermer 1997:26)? Are any of your beliefs based on pseudoscience?

But does it matter if we base our understanding of the social (and natural) worlds on the results of disciplined, documented tests of ideas in relation to empirical evidence—the scientific approach? Take a look at Figure 1.7. It shows that counties with a higher percentage of people fully vaccinated had lower rates of hospitalization for COVID-19 in early August 2021. The data

FIGURE 1.7 ■ Hospital COVID-19 Admissions and Vaccination Rates by County



Source: Harris, Jeffrey E. 2021. "COVID-19 Incidence and Hospitalization Rates Were Inversely Related to Vaccination Coverage Among the 112 Most Populous Counties in the United States." *Health Policy and Technology*, p. 3. https://pubmed.ncbi.nlm.nih.gov/34868833.

comes from a site managed by the U.S. Department of Health & Human Services (healthdata.gov) and was analyzed by a professor and medical doctor at the Massachusetts Institute of Technology. The article describes the data and methods of the statistical analysis in detail. Of course, it is only one article that presents just one slice of the picture about COVID-19 and vaccination and the author notes various other limitations, but it provides a clear foundation for the conclusion "that vaccination substantially reduces disease severity" (Harris 2021:8–9).

By the time you finish your study of how to investigate the social world, I hope you will appreciate the value of using the methods of science to understand the problems of society (and maybe even to improve your own health).

TYPES OF SOCIAL RESEARCH

Whatever the motives, there are four types of social research projects. This section illustrates each type with projects from the large body of research about various aspects of social ties.

Descriptive Research

Defining and describing social phenomena of interest is a part of almost any research investigation, but **descriptive research** is often the primary focus of the first research about some issue. Descriptive questions asked in research on social ties have included the following: What is the level of particular types of social ties in America (McPherson et al. 2006)? How has the frequency of different forms of social contact changed over time in Australia (Patulny and Seaman 2017)? What social and cultural patterns characterize disadvantaged neighborhoods (Harding 2007)? Measurement (the topic of Chapter 4) and sampling (Chapter 5) are central concerns in descriptive research. Survey research (Chapter 8) is often used for descriptive purposes. Some comparative research also has a descriptive purpose (Chapter 15).

Example: What is the extent of vaccine hesitancy?

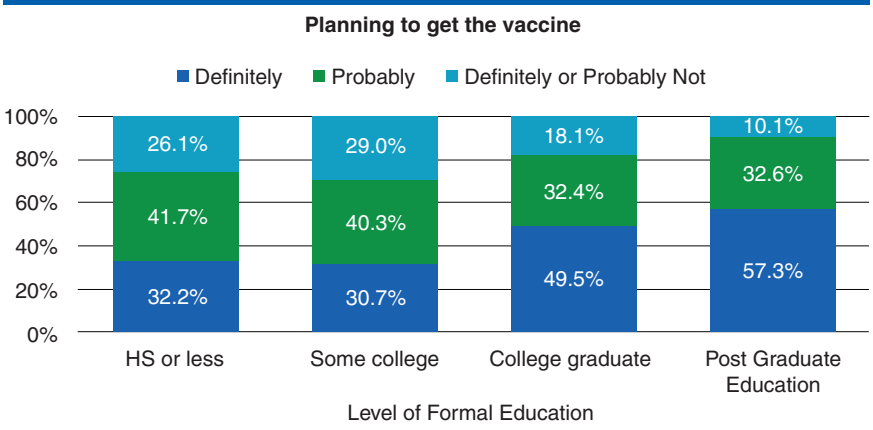
In a report based on our National Science Foundation-funded survey of Boston residents during the pandemic, Lee Hargraves, director of the University of Massachusetts Boston's Center for Survey Research and others on our team (2021) described the extent of vaccine hesitancy and its variation across racial and ethnic groups. Our survey of Boston adults included the following question:

If a vaccine against the coronavirus becomes available, do you plan to get vaccinated?

- Definitely
- Probably
- Probably not
- Definitely not

We found that, in Summer 2020, one in five Boston adults did *not* plan to get vaccinated. Among Black respondents, this fraction rose to almost half, compared to one quarter of Latinx respondents and just one in ten white and Asian/Pacific Islander respondents. As indicated in Figure 1.8, Hargraves et al. (2021) also found that vaccine hesitancy declined dramatically among college graduates and those with advanced degrees.

FIGURE 1.8 ■ Vaccine Hesitancy by Education



Source: Hargraves et al. 2021. "Living in Boston During COVID-19: Vaccination Planning and Hesitancy." Report #5. NSF-Report 5 Vaccination Hesitation (10JAN2021) v2.4 (umb.edu).

Exploratory Research

Exploratory research seeks to find out how people get along in the setting under question, what meanings they give to their actions, and what issues concern them. The goal is to learn "What is going on here?" and to investigate social phenomena without explicit expectations. This purpose is associated with the use of methods that capture large amounts of relatively unstructured information or that take a field of inquiry in a new direction. For example, researchers investigating social ties occurring through the Internet have had to reexamine the meaning of "community," asking whether cyberspace interactions can constitute a community that is seen as "real and essential" to participants (Fox and Roberts 1999:644). "How is identity—true or counterfeit—established in online communities?" asked Peter Kollock and Marc Smith (1999:9). Exploratory research such as this frequently involves qualitative methods, which are the focus of Chapters 10 and 11, as well as special sections in many other chapters.

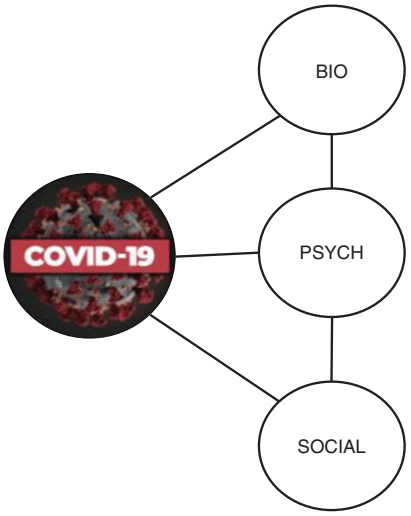
Example: How did young people experience the early months of the pandemic?

Portuguese social scientists Cátia Branquinho et. al (2020) were impressed with the dramatic impact of the pandemic on adolescents and wondered about teens' thoughts and experiences.

Branquinho and her colleagues decided to explore this question by conducting a survey in April and May 2020 that included eight open-response questions. Six hundred seventeen youth returned completed questionnaires. Using a systematic process, the researchers identified comments about biological, psychological, and social issues, with many negative consequences in these areas and some positive impacts (see Figure 1.9). For example, in the social area, one student said,

> Being in the 12th year, I am missing many moments that I waited for this year, whether be the finalists' trip or the ball or just being able to say goodbye to the secondary as I thought (Branquinho et al. 2020:2748).

FIGURE 1.9 ■ Types of Impacts of the Pandemic on Students



- **Physical symptoms** (- increased headache and muscle pain episodes)

- **Well-being** (+ happier times with school closed and online classes; more time for pleasurable activities, as physical exercise; — absence of a routine, with low productivity; changes in sleep patterns, food and exercise; less financial capacity, more homework)
- **Personal growth** (+ opportunity to set goals, and personal and training enrichment)
- **Psychological symptoms** (+ relaxation and less tiredness; - depression, anxiety, loneliness, irritability, etc.; greater technological dependence and substance use)

- **Relationships** (+ will help to understand the importance of relationships and strengthen friendships; - loss of contacts, decreased future physical contact, less confidence, decreased social skills; more family discussions and disruptions)
- **Life events** (loss of important moments)

Source: Branquinho, Cátia, Colette Kelly, Lourdes C. Arevalo, Anabela Santos, and Margarida Gaspar de Matos. 2020. "'Hey, We Also Have Something to Say': A Qualitative Study of Portuguese Adolescents' and Young People's Experiences Under COVID-19." *Journal of Community Psychology* 48(8):2744.

Branquinho et al. (2020) concluded that one impact of the pandemic was loss of social competencies due to less contact with others and missing key events like proms and trips.

Explanatory Research

Many philosophers of science consider explanation the premier goal of any science. **Explanatory research** seeks to identify the causes and effects of social phenomena and to predict how one phenomenon will change or vary in response to variation in some other phenomenon. I focus on ways of identifying causal effects in Chapter 6. Explanatory research often involves experiments (see Chapter 7) or surveys (see Chapter 8), both of which are most likely to use quantitative methods.

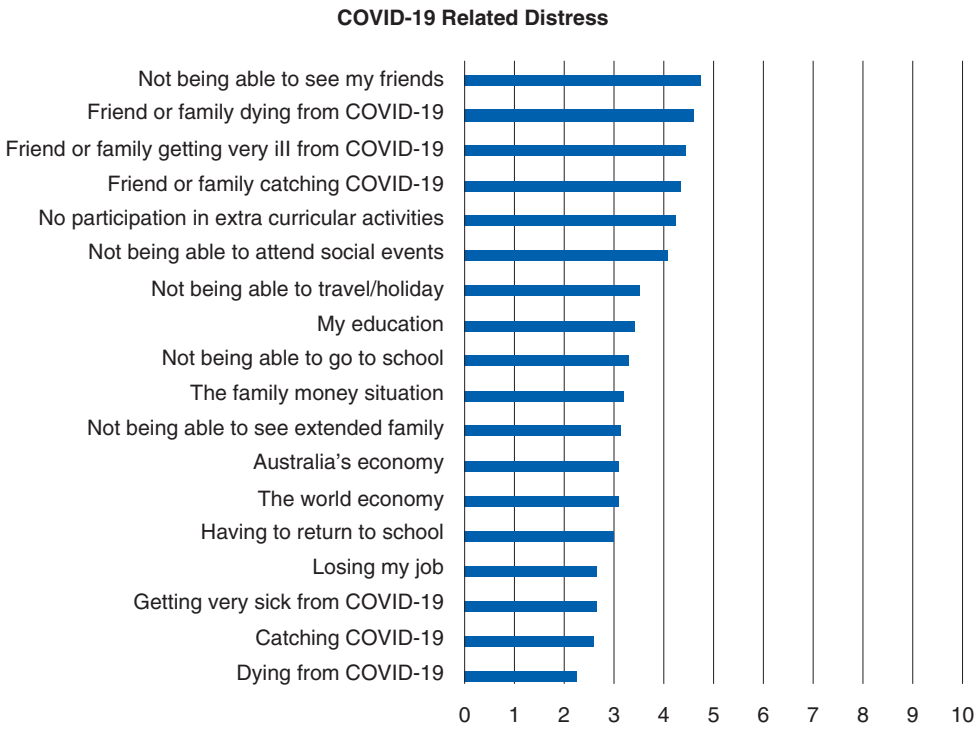
Example: What effect did the pandemic have on adolescents' mental health?

Natasha Magson et al. (2021) at Marquarie University in Sydney, Australia, sought to understand how the pandemic affected adolescents' mental health, why this effect occurred, and which factors increased or decreased mental health difficulties. For this purpose, they surveyed 248 adolescents before the pandemic and then two months after the Australian government-imposed restrictions after the pandemic began. The surveys included questions about anxiety, distress, life satisfaction, disruptions in schooling, media exposure, interpersonal conflict, and social connectedness.

Magson and her coauthors (2021) found that anxiety and distress had increased and life satisfaction had declined in the early months of the pandemic, although only by modest

amounts. The declines were more pronounced with girls than boys. The researchers also found that adolescents who felt more socially connected maintained better mental health, and those who followed the rules more by staying at home reported more life satisfaction. The researchers concluded that disruption to social connections caused more distress among adolescents than the virus itself (see Figure 1.10).

FIGURE 1.10 ■ Factors Contributing to COVID-19 Related Distress Among Adolescents


COVID-19 Related Distress
Not being able to see my friends
Friend or family dying from COVID-19
Friend or family getting very ill from COVID-19
Friend or family catching COVID-19
No participation in extra curricular activities
Not being able to attend social events
Not being able to travel/holiday
My education
Not being able to go to school
The family money situation
Not being able to see extended family
Australia's economy
The world economy
Having to return to school
Losing my job
Getting very sick from COVID-19
Catching COVID-19
Dying from COVID-19
0 1 2 3 4 5 6 7 8 9 10

Source: Magson, Natasha R., Justin Y. A. Freeman, Ronald M. Rapee, Cele E. Richardson, Ella L. Oar, and Jasmine Fardouly. 2021. "Risk and Protective Factors for Prospective Changes in Adolescent Mental Health During the COVID-19 Pandemic." *Journal of Youth and Adolescence* 50(1):44–57.

Evaluation Research

Evaluation research seeks to determine the effects of programs, policies, or other efforts to affect social patterns, whether by government agencies, private nonprofits, or for-profit businesses. This is a type of explanatory research because it deals with cause and effect, but it differs from other forms of explanatory research because evaluation research focuses on one type of cause: programs, policies, and other conscious efforts to create change (Lewis-Beck, Bryman, and Liao 2004:337). This focus raises some issues that are not relevant in other types of explanatory research, as explained in more detail in Chapter 13.

Concern regarding the potential impact of alternative policies regarding the Internet provided an impetus for new evaluation research.

Example: Do lockdowns during a pandemic change trust in government and science?

Covid-19 Related Distress

Strict government-imposed lockdowns to limit the spread of the coronavirus generated controversy and often halted regular social life as the pandemic spread around the globe. In the Netherlands, Oude Groeniger et al. (2021) evaluated the impact of the Dutch lockdown in mid-March 2020. Survey responses in the ongoing Longitudinal Internet Studies for the Social Sciences (LISS) allowed them to compare levels of trust before and after the imposition of the lockdown.

February 27, 2020, was the day of the first positive case of COVID-19 reported in the Netherlands. By March 12, the number of cases was increasing rapidly, and the Dutch prime minister, healthcare minister, and the government's scientific advisory board announced a lockdown, including restrictions on public gatherings, limitations on visiting, work-from-home arrangements when possible, and a stay-at-home order for those with symptoms.Groeniger and his coauthors (2021) found that trust in government was declining before the lockdown, but after it was imposed, trust in government rose by 18%. Trust in science had been greater to begin with but still rose by 6% after the lockdown began.

STRENGTHS AND LIMITATIONS OF SOCIAL RESEARCH

Using social scientific research methods to develop answers to questions about the social world reduces the likelihood of making everyday errors in reasoning. The various projects we have reviewed in this chapter illustrate this point:

- A clear definition of the population of interest increased the researchers' ability to draw conclusions without overgeneralizing findings to groups to which they did not apply. Use of a data set based on a broad, representative sample of the population enabled the Pew researchers (2020) Hargraves et al. (2021) and Groeniger et al. (2021) to describe attitudes during the pandemic in specific populations—adults in the United Staes, in Boston, or the Netherlands, respectively—rather than among some unknown set of their friends or acquaintances. The researchers' careful review of their sample design and the representativeness of the samples they obtained helps prevent overgeneralization to groups that were not actually studied.
- The use of surveys in each of these studies in which each respondent was asked the same set of questions reduced the risk of selective or inaccurate observation, as did careful attention to a range of measurement issues.
- The risk of illogical reasoning was reduced by carefully describing each stage of the research, clearly presenting the findings, and carefully testing the bases for cause-and-effect conclusions. For example, Magson et al. (2021) carefully compared the role of different factors that might have increased and decreased adolescents' mental health during the pandemic.
- Resistance to change was overcome in these studies by adhering to elements of the scientific method. For example, Magson et al. (2021) discussed possible alternative explanations for their findings (falsifiability), while Groeniger et al. (2021) compared

findings across different groups to determine whether more complex explanations were needed of behavioral change (simplicity). The publications by all the researchers help other researchers critique and learn from their findings (community) as well as inform the general public.

Nevertheless, I would be less than honest if I implied that we enter the realm of pure truth and light when we conduct social research or when we rely solely on the best available social research. Research always has some limitations and some flaws (as does any human endeavor), and our findings are always subject to differing interpretations. Social research permits us to see more, to observe with fewer distortions, and to describe more clearly to others what our opinions are based on, but it will not settle all arguments. Others will always have differing opinions, and some of those others will be social scientists who have conducted their own studies and drawn different conclusions.

In this way, one research study often leads to another, and another, each one improving on previous research or examining a research question from a somewhat different angle. Part of becoming a good social researcher involves learning to evaluate critically each research study and to weigh carefully the entire body of research about a research question before coming to a conclusion. And we have to keep an open mind about alternative interpretations and the possibility of new discoveries. The social phenomena we study are often complex, so we must consider this complexity when we choose methods to study social phenomena and when we interpret the results of these studies.

ALTERNATIVE RESEARCH ORIENTATIONS

In addition to deciding on the type of research they will conduct, social researchers also must choose among several alternative orientations to research.

Quantitative and/or Qualitative Methods

Different research methods provide different perspectives on social phenomena and so have different strengths and weaknesses. The most general distinction is between quantitative and qualitative research methods. Groeniger et al. (2021) analyzed the number of different responses to the same survey questions; we say this study used **quantitative methods**. Numerical data were also used in the descriptive survey about reactions to the pandemic (Pew 2021), as well as in the Hargraves et al. (2021) study about vaccine hesitancy and the Magson et al. (2021) research on adolescent mental health. In contrast, Branquinho et al. (2020) and Kee (2021) read and categorized text or written comments. Because the researchers focused on written or spoken text and its meaning, and did not attempt to quantify what they were studying, we say they used **qualitative methods**.

Quantitative methods are most often used when the motives for research are explanation, description, or evaluation. Exploration is more often—although by no means always—the motive for using qualitative methods. I highlight several other differences between quantitative and qualitative methods in Chapter 2. Chapters 10 and 11 present qualitative methods in much more detail.

Important as it is, I don't want to place too much emphasis on the distinction between quantitative and qualitative orientations or methods. Social scientists often combine these methods to enrich their understanding of the social world (Campbell and Russo 1999:141).

Philosophical Perspectives

Your preferences for particular research methods will be shaped in part by your general assumptions about how the social world can best be investigated—by your social research philosophy. The scientific approach reflects the belief that there is an objective reality apart from the perceptions of those who observe it. This is the philosophy traditionally associated with natural science and with the belief that scientists must be objective and unbiased to see reality clearly (Weber 1949:72). **Positivism** asserts that a well-designed test of a specific prediction—for example, the prediction that social ties decrease among those who use the Internet more—can move us closer to understanding actual social processes. Quantitative researchers are often guided by a positivist philosophy.

Postpositivism is a philosophy that is closely related to positivism because it also assumes an external, objective reality, but postpositivists acknowledge the complexity of this reality and the limitations and biases of the scientists who study it (Guba and Lincoln 1994:109–111). For example, postpositivists may worry that researchers who are heavy computer users themselves will be biased in favor of finding positive social effects of computer use. As a result of concerns such as this, postpositivists do not think we can ever be sure that scientific methods allow us to perceive objective reality as it actually is. Instead, they believe the goal of science is to achieve **intersubjective agreement** among scientists about the nature of reality (Wallace 1983:461). We can be more confident in the conclusions of the community of social researchers than in those of any individual social scientist (Campbell and Russo 1999:144).

Interpretivism is a research philosophy that emphasizes the importance of understanding subjective meanings people give to reality. Unlike positivism and postpositivism, interpretivism does not assume that social processes can be identified objectively.

Some qualitative researchers are guided by an interpretivist philosophy. You will learn more about different philosophical orientations in Chapter 10.

Basic Science or Applied Research

The effort to figure out what the world is like and why it works as it does—academic motivations—is the goal of **basic science** (Hammersley 2008:50). The Magson et al. (2021) study is a good example.

Social research may also have more immediate, practical concerns. Evaluation research like that conducted by Groeniger et al. (2021) on the effect of the lockdown by the Dutch government seeks to determine the impact of one or more programs or policies. This knowledge can then lead to practical changes, such as encouraging the use of lockdowns or alternatives to them. Evaluation research and other social research projects motivated by practical concerns are termed **applied research**.

Whether you think you would prefer a basic or applied orientation in social research, you have lots of company.

The Role of Values

The positivist and postpositivist philosophies consider value considerations to be beyond the scope of science: "An empirical science cannot tell anyone what he should do—but rather what he can do—and under certain circumstances—what he wishes to do" (Weber 1949:54). The idea is that developing valid knowledge about how society *is* organized or how we live our lives does not tell us how society *should* be organized or how we *should* live our lives. The determination of empirical facts should be a separate process from the evaluation of these facts as satisfactory or

unsatisfactory (Weber 1949:11). The idea is not to ignore value considerations but to hold them in abeyance during the research project until results are published.

There has always been tension between this "value-free" orientation to social research and a more "value-conscious" or even activist approach. In the 19th century, sociologist Lester Frank Ward (who subsequently became the American Sociological Society's first president) argued that "the real object of science is to benefit man. A science which fails to do this, however agreeable its study, is lifeless" (Ward 1897:xxvii). However, the American Sociological Society president in 1929, William Fielding Ogburn, urged the value-free approach: "Sociology as a science is not interested in making the world a better place to live. . . . Science is interested directly in one thing only, to wit, discovering new knowledge" (Ogburn 1930:300–301). Does one approach make more sense to you?

By the time you finish *Investigating the Social World*, I know you'll have a good understanding of the difference between these orientations, but I can't predict whether you'll decide one is preferable. Maybe you'll conclude they each have some merit.

Whether you plan to conduct your own research projects, read others' research reports, or just think about and act in the social world, recognizing the strengths and limitations of specific research methods and different approaches to social research will give you greater confidence in your own opinions, improve your ability to evaluate others' opinions, and encourage you to refine your questions, answers, and methods of inquiry.

CONCLUSIONS

I hope this first chapter has given you an idea of what to expect from the rest of the book. My aim is to introduce you to social research methods by describing what social scientists have learned about the social world as well as how they have learned it. The substance of social science is inevitably more interesting than its methods, but the methods become more interesting when they're linked to substantive investigations. I have focused attention in this chapter on research about social ties; in subsequent chapters, I introduce research examples from other areas.

Investigating the Social World is organized into four sections. The first section, Foundations for Social Research, includes the introduction in Chapter 1, and then an overview of the research process in Chapter 2 and an introduction to issues in research ethics and writing research proposals in Chapter 3. In Chapter 2, I review how social scientists specify research questions for investigation, how they orient themselves to those questions with social theories, and how they review related prior research. Most of the chapter focuses on the steps involved in the overall research process and the criteria that researchers use to assess the quality of their answers to the original research questions. Several studies of police–citizen interaction illustrate the research process in Chapter 2. Chapter 3, on research ethics and research proposals, completes the foundation for our study of social research. I emphasize in this chapter the importance of ethical treatment of human subjects in research and identify the problems caused by projects that have not adhered to this standard. I also introduce in this chapter the process of writing research proposals, which I then continue in the end-of-chapter exercises throughout the book. In actual research projects, submission of a research proposal to an institutional review board for the protection of human subjects is often the final step in laying the foundation for a research project.

The second section, Fundamentals of Social Research, presents methods for conceptualization and measurement, sampling, and causation and other elements of research design that must be considered in any social research project. In Chapter 4, I discuss the concepts we use to think

about the social world and the measures we use to collect data about those concepts. This chapter begins with the example of research on race and racism, but you will find throughout this chapter a range of examples from contemporary research. In Chapter 5, I use controversies over the 2020 Census and challenges of recent political polling to identify challenges in achieving generalizable results in research. In Chapter 6, I use research on violence to illustrate how to design research to answer such causal research questions as "What causes violence?" I also explain in this chapter the decisions social researchers must make about two research design issues that affect our ability to draw causal conclusions: (1) whether to use groups or individuals as units of analysis and (2) whether to use a cross-sectional or longitudinal research design.

The third section, Basic Social Research Designs, introduces the three primary methods of data collection and related methods of data analysis. Experimental studies, the subject of Chapter 7, focus attention on testing causal effects and are used often by social psychologists, psychologists, and policy evaluation researchers. Survey research is the most common method of data collection in sociology, so in Chapter 8, I describe the different types of surveys and explain how researchers design survey questions. I highlight in this chapter the ways in which the Internet and cell phones are changing the nature of survey research and use for some examples a survey on the impact of the pandemic. Chapter 9, on quantitative data analysis, introduces the statistics used to analyze data collected with experimental and survey designs. Chapter 9 is not a substitute for an entire course in statistics, but it provides the basic tools you can use to answer most research questions. To make this chapter realistic, I walk you through an analysis of quantitative data on voting in the 2020 presidential election. You can replicate this analysis with data on the book's study site (if you have access to the SPSS statistical analysis program or an alternative statistical package, including some that make it possible to analyze data online). You can also learn more about statistics with the SPSS exercises at the end of most chapters and with the study site.

Qualitative methods have long been the method of choice in anthropology, but they also have a long tradition in American sociology and have become the favored method of many social researchers around the world. Chapter 10 shows how qualitative techniques can uncover aspects of the social world that we are likely to miss in experiments and surveys and can sometimes result in a different perspective on social processes. Chapter 11 then focuses on the logic and procedures of analyzing qualitative data. In these chapters, you will learn about research on work organizations, psychological distress, gender roles, classroom behavior, and disasters such as Hurricane Katrina.

The fourth section, Complex Social Research Designs, presents research designs that can involve combinations of one or more of the basic research designs. By the time you read Chapter 12, you should be convinced of the value of using different methods to help us understand different aspects of the social world. Chapter 12 takes this basic insight a few steps further by introducing the use of "mixed methods." This increasingly popular approach to research design involves a careful plan for combining qualitative and quantitative methods in a research project. Evaluation research, the subject of Chapter 13, is conducted to identify the impact of social programs or to clarify social processes involving such programs. Evaluation research often uses experimental methods, but survey research and qualitative methods can also be helpful in evaluation research projects. Chapter14 reviews the methods of secondary data analysis and the related approach that has come to be known as "Big Data." In this chapter, you will learn how to obtain previously collected data and to investigate important social issues such as differences between urban neighborhoods.

Historical and comparative methods, the subject of Chapter 15, may involve either quantitative or qualitative methods that are used to compare societies and groups at one point in time and to analyze their development over time. We will see how these different approaches have been used to learn about political change in transitional societies. I also explain the method of content analysis in this chapter; it can be used in historical and comparative research and provides another way to investigate social processes in an unobtrusive way.

Plan to read Chapter 16 carefully. Our research efforts are only as good as the attention given to our research reports, so my primary focus in this chapter is on writing research reports. I also present means for enhancing graphic displays to communicate quantitative results more effectively in research reports. In addition, I introduce meta-analysis—a statistical technique for assessing many research studies about a particular research question. By the end of the chapter, you should have a broader perspective on how research methods can improve understanding of the social world.

Each chapter ends with several helpful learning tools. Lists of learning goals, chapter highlights and key terms will help you review the ideas that have been discussed. Discussion questions will help you and your classmates apply and deepen your knowledge. Statistical exercises will introduce you to the process and logic of quantitative analysis and demonstrate its relevance to most of the chapter topics.

KEY TERMS

Applied research (p. 24)
Basic science (p. 24)
Descriptive research (p. 18)
Evaluation research (p. 21)
Explanatory research (p. 20)
Exploratory research (p. 19)
Illogical reasoning (p. 12)
Inaccurate observation (p. 10)
Interpretivism (p. 24)
Intersubjective agreement (p. 24)
Overgeneralization (p. 11)
Positivism (p. 24)
Postpositivism (p. 24)
Pseudoscience (p. 17)
Qualitative methods (p. 23)
Quantitative methods (p. 23)
Resistance to change (p. 12)
Science (p. 14)
Selective observation (p. 9)
Social science (p. 16)

HIGHLIGHTS

LO 1.1 Describe three common types of motivations for social research.

Three common types of motivations for social research are policy, academic, and personal motivations.

- Policy motivations include when public and private entities (e.g., government agencies) seek to better understand social ties to inform policy recommendations.
- Academic motivations reflect social scientists' desires to understand how the social world works.
- Personal motivations refer to social research conducted to address problems in the social world to which one has been exposed.

LO 1.2 Illustrate each of the four common errors involved in learning about the social world with a hypothetical example.

We can avoid making errors in four processes involved in learning about the social world in the following ways:

- Observing—not participating in selective observation (choosing to look only at things that are in line with our preferences or beliefs) or inaccurate observation (an observation based on faulty perceptions of empirical reality).
- Generalizing—avoiding overgeneralization (when we unjustifiably conclude that what is true for some cases is true for all cases).
- Reasoning—steer clear of illogical reasoning (when we prematurely jump to conclusions or argue based on invalid assumptions).
- Reevaluating—abstain from resistance to change (the reluctance to change our ideas in response to new information), perhaps related to ego-based commitments, excessive devotion to tradition, and uncritical agreement with authority.

LO 1.3 List components of the scientific approach.

The scientific approach involves investigating phenomena in the world by testing ideas about them with observations—empirical data—of those phenomena.

- Science uses logical, systematic, documented methods.
- Science investigates nature and natural processes.
- Science is the knowledge produced by these investigations.
- Science is an ongoing, cumulative process.
- Science is empirical and objective (or intersubjective).

LO 1.4 Compare the four types of social research.

Social research can be descriptive, exploratory, explanatory, or evaluative—or some combination of these.

- Descriptive research: Research in which social phenomena are defined and described
- Exploratory research: Research that seeks to find out how people get along in the setting under question, what meanings they give to their actions, and what issues concern them
- Explanatory research: Research that seeks to identify causes and effects of social phenomena and to predict how one phenomenon will change or vary in response to variation in some other phenomenon
- Evaluation research: Research that describes or identifies the impact of social policies and programs

LO 1.5 Assess the strengths and limitations of social research.

- The strengths of social research include a reduced likelihood of making everyday errors in reasoning.
- Social research permits us to see more, to observe with fewer distortions, and to describe more clearly to others what our opinions are based on.
- The limitations of social research include the fact that social research findings are always subject to differing interpretations and opinions.
- Social research cannot resolve value questions or provide permanent, universally accepted answers.

LO 1.6 Contrast these pairs of alternatives in social research: positivism and postpositivism; quantitative methods and qualitative methods; basic research and applied research.

- Positivism is the belief shared by most scientists, that there is a reality that exists quite apart from our own perception of it, that it can be understood through observation, and that it follows general laws.
- Postpositivism is a philosophical view that modifies the positivist premise of an external, objective reality by recognizing its complexity, the limitations of human observers, and therefore the impossibility of developing more than a partial understanding of reality.
- Quantitative methods include data that capture variation in social life in either numbers or attributes that can be ordered by magnitude.
- Qualitative methods include data based on actual observations of social life as participants experience it that do not often have a direct numerical interpretation.
- Basic science research focuses on expanding knowledge and providing results to other researchers.
- Applied research seeks to have an impact on social practice and to share results with a wide audience.

DISCUSSION QUESTIONS

1. Select a social issue that interests you, such as Internet use or crime. List at least four of your beliefs about this phenomenon. Try to identify the sources of each of these beliefs.
2. Does the academic motivation to do the best possible job of understanding how the social world works conflict with policy or personal motivations? How could personal experiences with social isolation or with Internet use shape research motivations? In what ways might the goal of influencing policy about social relations shape a researcher's approach to this issue?
3. Pick a contemporary social issue of interest to you. Describe different approaches to research on this issue that would involve descriptive, exploratory, explanatory, and evaluative approaches.
4. Review the strengths of social research. How convinced are you about each of them at this point?
5. Review each of the research alternatives. Do you find yourself more attracted to a quantitative or a qualitative approach? To a positivist, postpositivist, or interpretivist philosophy? To doing research to contribute to basic knowledge or to shape social policy? What do you think about value freedom as a standard for science?

SPSS EXERCISES

1. What personal motivation might you have for studying the level of concern about the COVID-19 pandemic? What might motivate other people to conduct research on this topic? What policy and academic motives might be important?

2. After you download one of the GSS2020 files and save it in a directory, open the GSS2020 file. In the SPSS menu, click on File, then on Open and Data, and then on the name of the data file in the directory where it is saved. How many respondents are there in this subset of the complete GSS file? (Scroll down to the bottom of the data set in Data View.) How many variables were measured? (Scroll down to the bottom of the Variable View in SPSS.)

3. What would you estimate as the level of concern about the COVID-19 pandemic in the United States in 2020? Now for your first real research experience in this text: Describe the distribution of concern about the COVID-19 pandemic (WRYCOVID_2). Obtaining the relevant data is as simple as "a, b, c, d, e."
   - **a.** Click on Graphs.
   - **b.** Click on Legacy Dialogs > Bar.
   - **c.** Select "Simple" and "Summaries for groups of cases" under Data in Chart Area > Define.
   - **d.** Place the WRYCOVID_2 variable in the box below "Category Axis:" and select "% of cases" under "Bar Represent."
   - **e.** Click OK.

Now describe the distribution of concern about the COVID-19 pandemic. What percentage of the population was very concerned with the COVID-19 pandemic in 2020?

©iStockphoto.com/Josie Desmarais

2 THE PROCESS AND PROBLEMS OF SOCIAL RESEARCH

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

2.1 Appraise social research questions based on their feasibility, social importance, and scientific relevance

2.2 Compare the concepts of "theory" and "paradigm"

2.3 Devise a strategy for searching the literature and the web about a research question

2.4 Design a literature review for a research question

2.5 Assess whether a research project is designed to answer primarily a deductive or an inductive research question

2.6 Compare the four standards for social research: measurement validity, generalizability, causal validity, and authenticity

Michael Brown, Eric Garner, George Floyd. Their names and those of many other Black victims of police shootings highlight racial disparities in the U.S. criminal justice system. Over their lifetimes, we now know the risk of Black men being killed by police is about 2.5 times higher than it is for white men (Figure 2.1). But even as widespread protests erupted after the 2014 fatal shooting of Michael Brown in Ferguson, Missouri, "the absence of definitive official data" made it difficult to estimate the prevalence of police-involved deaths (Edwards, Lee, and Esposito 2019:16793). Then-FBI director James Comey called this lack of adequate data "unacceptable" and "embarrassing" (Peeples 2019). Distinguished criminologist Lawrence W. Sherman (2020:8) called for "reflective inquiry,"

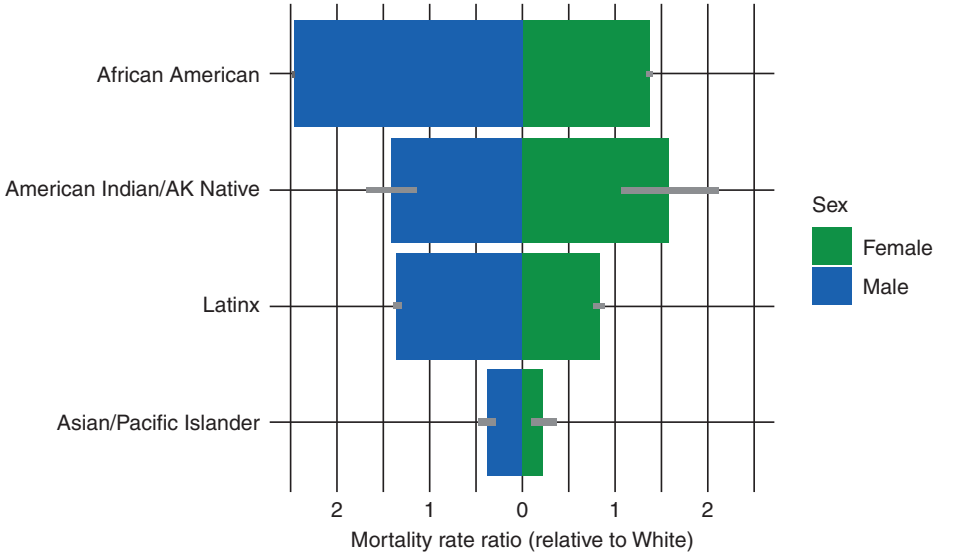
> Undertaken by a range of talents, disciplines, and professions, working together to ask now not just how each death might be prevented, but how patterns of death might be predicted and interrupted.

In the years since Ferguson, systematic data collection and many social science investigations have allowed credible estimates of the prevalence of police shootings, answered many questions about who is more likely to be shot by police, and tested various ways of reducing the likelihood of police shootings. The resulting research provides many compelling examples of social science in practice and more than a few indications of its limitations. I will discuss several of these research studies and other policing-related research in this chapter to illustrate how social scientists formulate questions for research, develop a foundation in theory and prior research, use different research strategies, and take each of the steps that together make up the research process (see Figure 2.2). The chapter also expands on the role of social theories in developing research questions and guiding research decisions. I will conclude by identifying the standards that guide this process. Appendix A lists questions to ask when reviewing a research article; I refer to these questions when I illustrate the article review process in this chapter.

You will find throughout the chapter examples of how social science research has improved understanding of police behavior and provided a foundation for effective change. By the chapter's end, you should be ready to formulate a research question about any issue, critique previous studies that addressed this question, and design a

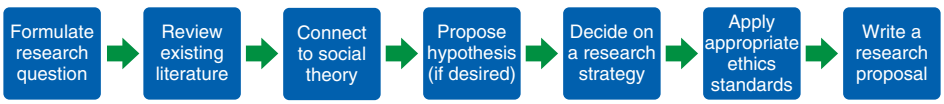
general strategy for answering the question. In the next chapter, you will learn how to review ethical issues and write a research proposal.

FIGURE 2.1 ■ Inequality in Lifetime Risk of Being Killed by Police Use of Force in the United States



Source: Edwards, Frank, Hedwig Lee, and Michael Esposito. 2019. "Risk of Being Killed by Police Use of Force in the United States by Age, Race–Ethnicity, and Sex." *Proceedings of the National Academy of Sciences of the United States of America* 116(34):16793–16798.

FIGURE 2.2 ■ Launching a Research Project



SOCIAL RESEARCH QUESTIONS

A **social research question** is a question about the social world that one seeks to answer through the collection and analysis of firsthand, verifiable, empirical data. It is not a question about who did what to whom but a question about people in groups, about general social processes, or about tendencies in community change, such as the following: What distinguishes police who have had records of unjustified shootings of civilians? Does community policing reduce the crime rate? What influences the likelihood of spouse abuse? How do people react to social isolation? So many research questions are possible that it is more of a challenge to specify what does *not* qualify as a social research question than to specify what does.

But that doesn't mean it is easy to specify a research question. In fact, formulating a good research question can be surprisingly difficult. We can break the process into three stages: (1) identifying one or more questions for study, (2) refining the questions, and then (3) evaluating the questions.

Identifying Social Research Questions

Social research questions may emerge from your own experience—from your "personal troubles"—as C. Wright Mills (1959) put it. One experience might be membership in a church, another could be victimization by crime, and yet another might be moving from a dorm to a sorority house. You may find yourself asking a question such as, "How do police shootings affect community trust in government?" "In what ways do people tend to benefit from church membership?" or "How do initiation procedures influence group commitment?" What other possible research questions can you develop based on your own experiences in the social world? Now is a good time to start thinking of research questions that interest you and others (Firebaugh 2008:2).

The research literature is often the best source for research questions. Every article or book will bring new questions to mind. Even if you're not feeling too creative when you read the literature, most research articles highlight unresolved issues and end with suggestions for additional research. For example, Stephen Wu (2021) concluded an article on the race of a city's police chief and the incidence of police officer fatal shootings by suggesting, "Further study of the channels through which the race of police leaders impacts the use of deadly force by officers would be a fruitful avenue for further research" (p. 416). A new study could focus on these mechanisms: Do Black police chiefs have a different management style? Do Black and white officers differ in their response to Black police chiefs? Do changes in behavior reflect changes in officer attitudes or supervisory sanctions? Any research article in a journal is likely to suggest other unresolved issues about the research question studied.

Many social scientists find the source of their research questions in social theory. Some researchers spend much of their careers conducting research intended to refine an answer to one research question that is critical for a particular social theory. For example, one theory explains social deviance as resulting from labels that people attach to some behaviors. A researcher could ask whether this "labeling theory" can explain some of the effects of arrest.

Finally, some research questions have very pragmatic sources. You may focus on a research question someone else posed because it seems to be to your advantage to do so. Some social scientists conduct research on specific questions posed by a funding source in what is termed an RFP, a request for proposals. (Sometimes the acronym RFA is used, meaning request for applications.) Others may find support from a government entity or foundation that supports research in specific areas of interest. Joscha Legewie's (2016) study of racial profiling and the disproportionate use of force was funded by a grant from the Russell Sage Foundation. Anthony Braga and his research collaborators (2020) received funding from the City of Boston and the Rappaport Institute for Greater Boston for their study of the effects of body-worn cameras. Or you may learn that the social workers in the homeless shelter where you volunteer need help with a survey to learn about client needs, which becomes the basis for another research question.

Refining Social Research Questions

It is even more challenging to focus on a problem of manageable size than it is to come up with an interesting question for research. We are often interested in much more than we can reasonably investigate with limited time and resources. In addition, researchers may worry about staking a research project (and thereby a grant or a grade) on a single problem, and so they may address several research questions at once. Also, it might seem risky to focus on a research question that may lead to results discrepant with our own cherished assumptions about the social world. The

prospective commitment of time and effort for some research questions may seem overwhelming, resulting in a certain degree of paralysis.

The best way to avoid these problems is to develop the research question gradually. Don't keep hoping that the perfect research question will just spring forth from your pen. Instead, develop a list of possible research questions as you go along. At the appropriate time, look through this list for the research questions that appear more than once. Narrow your list to the most interesting, most workable candidates. Revise your research questions and repeat this process as long as it helps improve your research question.

Evaluating Social Research Questions

In the third stage of selecting a research question, you evaluate the best candidate against the criteria for good social research questions: feasibility, given the time and resources available; social importance; and scientific relevance (King, Keohane, and Verba 1994).

Feasibility

You must be able to conduct any study within the time and resources available. If time is short, questions that involve long-term change may not be feasible. Another issue is to what people or groups we can expect to gain access. For example, observing social interaction in corporate boardrooms may be taboo. Next, you must consider whether you will have any additional resources, such as research funds or other researchers with whom to collaborate. Remember that there are severe limits on what one person can accomplish. However, you may be able to piggyback your research onto a larger research project. You also must consider the constraints faced because of your schedule, other commitments, and skill level, not to mention what is ethically defensible.

A study of police violence and citizen crime reporting by Matthew Desmond, Andrew V. Papachristos, and David S. Kirk (2016) shows how ambitious a social research question can be, even without major external funding. The key to feasibility in this case was the Milwaukee Police Department's agreement to provide the researchers with data from all 911 calls placed in Milwaukee for the year before and after the publicized beating of a Black man (Frank Jude) by off-duty Milwaukee police officers. Funding from Harvard University (Desmond's home institution) supported the time required to analyze data. Desmond, Papachristos, and Kirk note that other researchers could attempt to repeat their study in other locales that are willing to release administrative records. Of course, students in a research methods class may not be granted access to such data from the local police department, but you will learn in Chapter 14 that many datasets are available for analysis (including records of 911 calls in Boston!). When funding or data access limitations create barriers, you might alter your research question to make it more feasible. For example, you might ask "How do students at my university react to images of police shootings?"

Social Importance

Social research is not a simple undertaking, so it's hard to justify the expenditure of effort and resources unless you focus on a substantive area that is important. Besides, you need to feel motivated to carry out the study. Nonetheless, "importance" is relative, so for a class assignment, focusing on student reactions to dormitory rules or something similar might be important enough.

For most research undertakings, you should consider whether the research question is important to other people. Will an answer to the research question make a difference for society

or for social relations? Again, Desmond et al's (2016) study of police violence and citizen crime reporting is an exemplary case. But the social sciences are not wanting for important research questions. The June 2021 issue of the *American Sociological Review*—the journal that published the Desmond, Papachristos, and Kirk (2016) article—contained articles on politics and infant health, trauma and nationalism, selling student papers in Uganda, marriage and masculinity in 19 countries, and generosity. All these articles addressed research questions about important social issues, and all raised new questions for additional research.

The examples of good and poor research questions in Table 2.1 give you an idea of how to apply these criteria.

TABLE 2.1 ■ Evaluating Research Questions

Research Question	Quality	Explanation
Why are there wars?	Poor	Too general
Why is my uncle an alcoholic?	Poor	Too specific
Why are some people more prone to alcoholism than others?	Good	Many feasible ways to investigate
Is there a God?	Poor	Not answerable with empirical research
Are married men happier than those who are not married?	Poor	Abundant research already confirms this. Try a different angle
Does torture increase likelihood of confession to crimes?	Poor	Unethical to study by direct manipulation
Does group living increase housing retention for formerly homeless persons more than living in independent apartments?	Good, depending	Feasible for direct test if millions of dollars to fund residential alternatives

Source: Based on Firebaugh, Glenn. 2008. *Seven Rules for Social Research*. Princeton, NJ: Princeton University Press.

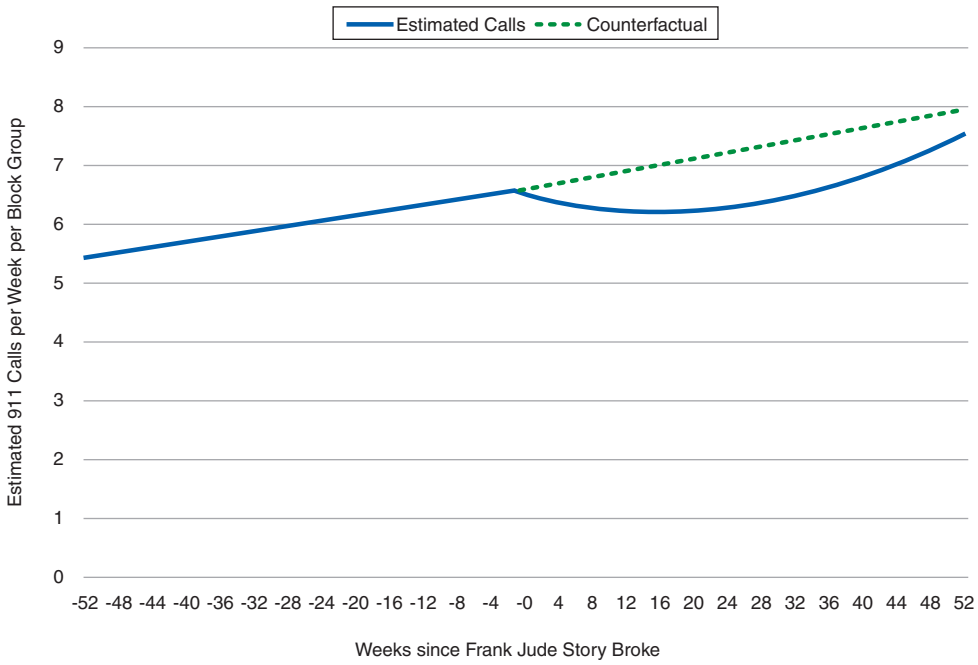
Scientific Relevance

Every research question should be grounded in the social science literature. Whether you formulate a research question because you have been stimulated by an academic article or because you want to investigate a current social problem, you should first turn to the social science literature to find out what has already been learned about this question. You can be sure that some prior study is relevant to any research question you can think of.

The study of police violence and citizen crime reporting was grounded in legal cynicism theory, which expects a lack of cooperation with police in underrepresented communities due to the belief that police lack interest in assisting residents (Baumer 2002; Kirk and Papachristos 2011). In the Milwaukee study, the theoretically based prediction was that publicity about Frank Jude's beating would increase legal cynicism in the underrepresented community and so result in fewer 911 calls for police assistance. This is exactly what the researchers found: 911 calls to the

police dropped immediately after the story was publicized (see Figure 2.3). The research now can be used by other social scientists seeking to develop and test legal cynicism theory.

FIGURE 2.3 ■ Estimated Number of Weekly 911 Calls Reporting Crime Before and After the Frank Jude Story



Source: Desmond, Matthew, Andrew V. Papachristos, and David S. Kirk. 2016. "Police Violence and Citizen Crime Reporting in the Black Community." *American Sociological Review* 81(5):857–76.

What if there is already a lot of convincing research published about your proposed research question? Can you focus the question on a new population or setting? Is early intervention with social services more effective in reducing violence among juveniles than among adults? Do men like using Facebook as much as women do? You get the idea.

SOCIAL THEORIES

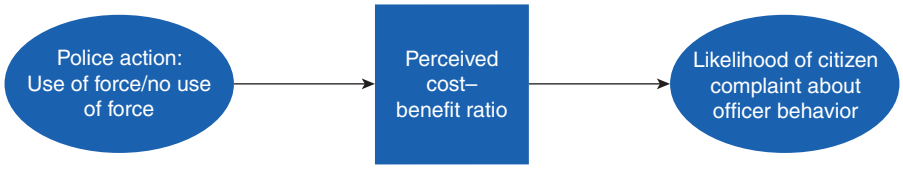
Police behavior does not exist in a vacuum, set apart from the rest of the social world. We can understand behaviors and orientations better if we consider how they reflect broader social patterns. Does the behavior of some police officers reflect their conceptions of masculinity? Do communities vary in their level of social cohesion and physical safety? Our answers to general questions such as these will help shape the research questions we ask and the methods we use.

Although everyone has general notions about "how things work," "what people are like," and so on, social scientists draw on more formal sets of general ideas—social theories—to guide their research (Collins 1994). A **theory** is a logically interrelated set of propositions that helps us make sense of many interrelated phenomena and predict behavior or attitudes that are likely to occur when certain conditions are met. Theory helps social scientists decide which questions are important to ask about the social world and which are just trivial pursuits. Theory focuses a spotlight on the particular features of the social world where we should look to get answers for

these questions, how these features are related to each other, and what features can be ignored. Building and evaluating theory is therefore one of the most important objectives of social science.

Anthony Braga and his collaborators (2020) at Northeastern University designed their study of the effect of body-worn cameras (BWCs) on police–citizen encounters to test predictions from deterrence theory. Deterrence theory is itself based on a broader perspective that is termed **rational choice theory**. Rational choice theory assumes that people's behavior is shaped by practical cost–benefit calculations (Coleman 1990:14). *Deterrence theory* applies rational choice theory to crime and punishment (Lempert and Sanders 1986:86–87), proposing that offenders are less likely to commit crimes if they believe the costs of illegal acts will exceed their benefits (Braga et al. 2020). Crime "doesn't pay" (as much) when people are aware of the costs of punishment (see Figure 2.4). Braga et al.'s (2020) prediction from deterrence theory about police behavior was that police officers should be less likely to engage in illegal behavior if they know their actions are being recorded and therefore are more likely to result in the punishment of the officer. A different theory, public self-awareness theory also led Braga et al. (2020) to predict more compliance with social norms by police wearing a camera. The theory expects that when people are being observed (a consequence of a BWC), they are reminded of social norms and are more likely to adjust their behavior to comply with those norms.

FIGURE 2.4 ■ Rational Choice Theory Prediction


Police action: Use of force/no use of force
Perceived cost–benefit ratio
Likelihood of citizen complaint about officer behavior

Do these concepts interest you? Do these propositions strike you as reasonable ones? If so, you might join a long list of researchers who have attempted to test, extend, and modify various aspects of rational choice theory.

Some sociologists attempt to understand the social world by investigating the meaning people attach to their interactions. These researchers focus on the symbolic nature of social interaction—how social interaction conveys meaning and promotes socialization. Herbert Blumer developed these ideas into **symbolic interaction theory** (Turner, Beeghley, and Powers 1995:460).

Labeling theory uses a symbolic interactionist approach to explain deviance as an "offender's" reaction to the application of rules and sanctions (Becker 1963:9; Scull 1988:678). Can you see how researchers like Braga et al. (2020) could use labeling theory to make predictions about the effects of body-worn cameras? It would involve a somewhat different rationale than self-awareness theory for explaining the effect of BWCs. Police officers might want to avoid being labeled as rule violators after being "caught on camera"; so too could those who might be subject to arrest. Projecting camera images of deviant behavior in the courtroom could also increase the likelihood of labeling officers and those arrested as deviant. In an earlier study of the police response to domestic violence complaints, Sherman and Berk (1984) recognized that a labeling process might influence offenders' responses to arrest. Once the offender is labeled as a deviant by undergoing arrest, other people treat the offender as deviant, and they are then more likely to act in a way that is consistent with the label *deviant*. Ironically, the act of punishment stimulates more of the very behavior it was intended to eliminate.

Do you find yourself thinking of some interesting research foci when you read about this labeling theory of deviance? If so, consider developing your knowledge of symbolic interaction theory and using it as a guide in your research.

Conflict theory focuses on basic conflicts between different social groups in society and how groups attempt to exercise domination to their own benefit (Collins 1994:47). The theory has its origins in Karl Marx and Friedrich Engels's (1961:13–16) focus on social classes as the key groupings in society and their belief that conflict between social classes was the "engine" of social change.

Although different versions of conflict theory emphasize different bases for conflict, they focus attention on the conflicting interests of groups rather than on individuals' concerns with maximizing their self-interest. As applied to crime, conflict theory suggests that laws and the criminal justice system are tools of the upper classes to maintain their dominance over the lower classes.

Tammy Rinehart Kochel (2019) turned to conflict theory to explain the impact of "Ferguson" (the fatal police shooting of Michael Brown) on residents' views of the police. Conflict theory predicts that people in a disadvantaged group within a political system will tend to be distrustful of representatives of that system—such as police (Kochel 2019). Findings of Kochel's before-and-after survey of the residents in the county in which Ferguson is located were consistent with this prediction: Black residents' trust in police declined after the shooting, while white residents' level of trust in police did not change.

Do these concepts strike a responsive chord with you? Can you think of instances in which propositions of conflict theory might help explain social change?

French social theorist Émile Durkheim used a very different theory, **functionalism**, to explain forms of deviance from societal norms. Writing during the period of rapid social change in Europe at the dawn of the 20th century, Durkheim (1893/2014) was concerned with the strength of social bonds in society. He posited that traditional social bonds based on similarity between people were being replaced by social bonds based on interdependence between people performing different social roles. For example, urban dwellers needed farmers to grow their food, truckers to bring the crops to market, merchants to arrange the sale of the crops, butchers to prepare meat, cobblers to make shoes, and so forth. Durkheim (1893/2014) termed social bonds based in this way on interdependence as *organic solidarity* (bringing to mind the interdependence of different organs in the body). Crime is explained by functionalists as occurring because it is functional for society to delimit the boundaries around acceptable behavior.

As a social researcher, you may work with one of these theories, seeking to extend it, challenge it, or specify it. You may test alternative implications of the different theories against each other. If you're feeling ambitious, you may even seek to combine some aspects of the different perspectives. Maybe you'll come up with a different theoretical perspective altogether. Or you may find you lose sight of the larger picture in the midst of a research project; after all, it is easier to focus on accumulating particular findings rather than considering how those findings fit into a more general understanding of the social world. But you'll find, in any area of research, developing an understanding of relevant theories will help you ask important questions, consider reasonable alternatives, and choose appropriate research procedures.

Scientific Paradigms

Scientific paradigms are sets of beliefs that guide scientific work in an area, including unquestioned presuppositions, accepted theories, and exemplary research findings. In his famous book on the history of science, *The Structure of Scientific Revolutions*, Thomas S. Kuhn (1970) argued

that most of the time, one scientific paradigm is accepted as the prevailing wisdom in a field, and scientists test ideas that make sense within that paradigm. They are conducting what Kuhn called **normal science**. Only after a large body of contrary evidence accumulates might there be a rapid shift to a new paradigm (Hammersley 2008:46).

Some people refer to conflict theory, functionalist theory, and symbolic interaction theory as alternative paradigms, although this stretches the meaning of *paradigm* a bit. In any case, many social scientists draw on more than one of these theories (and others) in their research and so reject the notion that they are truly incommensurable paradigms. We also should be sensitive to the insights that can be provided by examining social phenomena from different perspectives.

SEARCHING FOR PRIOR RESEARCH

How do you find prior research and theorizing on questions of interest? You may already know some of the relevant material from prior coursework or your independent reading, but that won't be enough. When you begin to focus on a research question, you need to search for academic articles focused on that question and its different components, with an emphasis on those based on research projects using methods like those you plan to use. If there have been no prior studies of the same exact research question on which you want to focus, you should seek articles from investigations of very similar research questions. Once you have located articles from prior research like the research you want to conduct, you may expand your search to include investigations about related topics or studies that used similar methods.

Searching the web can also identify research projects, unpublished reports, and other resources that may be useful as you explore your research question from different angles and seek related resources. This process requires a different approach than searching the academic literature, so I will treat it in its own section.

Searching the Literature

Conducting a thorough search of the research literature and then reviewing critically what you have found lays an essential foundation for any research project. Fortunately, much of this information can be identified online, without leaving your desktop or laptop, and an increasing number of published journal articles can be downloaded directly onto your own computer (depending on your own access privileges). But just because there's a lot available online doesn't mean you need to find it *all*. Keep in mind that your goal is to find relevant reports of prior research investigations.

The type of reports you should focus on are those that have been screened for quality through critique by other social scientists before publication. Scholarly journals, or *refereed journals* that publish *peer-reviewed articles*, manage this review process. Most often, editors of refereed journals send articles that authors submit to three or more other social scientists for anonymous review. Based on the reviewers' comments, the journal editor then decides whether to accept or reject the article or to invite the author to "revise and resubmit." This process results in the rejection of most articles (top journals such as the *American Sociological Review* or the *American Journal of Sociology* may reject about 90% of the articles submitted), and those that are ultimately accepted for publication typically have to be revised and resubmitted first. This helps ensure a much higher-quality standard, although journals vary in the rigor of their review standards, and, of course, different reviewers may be impressed by different types of articles; you thus always have to make your own judgment about article quality.

Newspaper and magazine articles may raise important issues or summarize social science research investigations, but they are not an acceptable source for understanding the research literature. The web offers much useful material, including research reports from government and other sources, sites that describe social programs, and even indexes of the published research literature. You may find copies of rating scales, reports from research in progress, papers that have been presented at professional conferences, and online discussions of related topics. Web search engines will also find academic journal articles you can access directly online, although usually for a fee. Most of the published research literature will be available to you online only if you go through the website of your college or university library. The library pays a fee to companies that provide online journals so you can retrieve this information without paying anything extra yourself. Of course, no library can afford to pay for every journal, so if you can't find a particular issue of a particular journal you need online, you will have to order the article you need through interlibrary loan or, if the hard copy of the journal is available, walk over to your library to read it.

As with any part of the research process, your method for searching the literature will affect the quality of your results. Your search method should include the following steps:

1. *Specify your research question.* Your research question should be neither so broad that hundreds of articles are judged relevant nor so narrow you miss important literature. "Is informal social control effective?" is probably too broad. "Does informal social control reduce rates of burglary in my town?" is probably too narrow. "Is informal social control more effective in reducing crime rates than policing?" provides about the right level of specificity.

2. *Identify appropriate bibliographic databases to search.* Sociological Abstracts or SocINDEX may meet many of your needs, but if you are studying a question about social factors in illness, you should also search in MEDLINE, the database for searching the medical literature. If your focus is about psychological issues, you'll also want to include a search in the online Psychological Abstracts database, PsycINFO or the version that also contains the full text of articles, PsycARTICLES. Search Criminal Justice Abstracts if your topic is in the area of criminology or criminal justice, or EconLit, if your topic might be addressed in the economic literature. A multidisciplinary bibliographic database like Academic OneFile or Academic Search Complete may be a good choice for some research questions. It will save you a lot of time if you ask a librarian to teach you the best techniques for retrieving the most relevant articles to answer your questions.

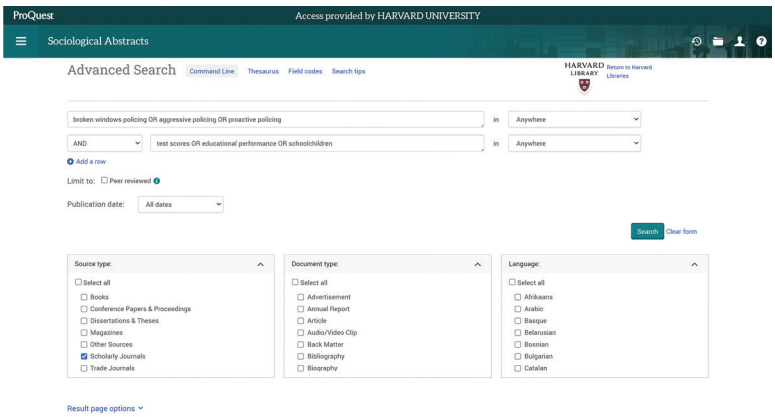
   To find articles that refer to a previous publication, such as Braga et al.'s (2018) early report on the effect of body-worn cameras, the Social Science Citation Index (SSCI) will be helpful. SSCI is an extremely useful tool for tracing the cumulative research in an area across the social sciences. SSCI has a unique "citation searching" feature that allows you to look up articles or books, see who else has cited them in their own work, and find out which articles and books have had the biggest impact in a field.

3. *Create a tentative list of search terms.* List the parts and subparts of your research question and any related issues you think are important: "informal social control," "police use of force," "body-worn cameras," and perhaps "trust in the police." It might help to start with one key article that has focused on your research question, identify the concepts in it, and search those concepts. You can then expand the search with more concepts in the articles you locate (Booth, Sutton, and Papaioannou 2016:115).

4. *Narrow your search.* The sheer number of references you find can be a problem. For example, searching for "social capital" (in quotation marks) in April 2022 resulted in 9,589 citations in SocINDEX with full text. Depending on the database you are working with and the purposes of your search, you may want to limit your search to English-language publications, to journal articles rather than conference papers or dissertations (both of which are more difficult to acquire), and to materials published in recent years. If your search yields too many citations, try specifying the search terms more precisely (e.g., "neighborhood social capital"). If you have not found much literature, try using more general or multiple terms (e.g., "social relations" or "social ties"). You may want to include the names of authors of major relevant studies or of the most important relevant journals in addition to the subject terms. Whatever terms you search first, don't consider your search complete until you have tried several different approaches and have seen how many articles you find. A search for "police shootings" in SocINDEX on April 9, 2022, yielded 220 hits; by adding "effects or impact or consequences or influence or outcomes" as required search terms (on a second search line, without quotation marks) and limiting the search to peer-reviewed articles published since 2010 (check search options for the database), the number of hits dropped to 62. So you may need to spend some time trying different combinations of search terms and search limiters. Also, consider trying your search in a different database or in several databases at once. When I redid my last SocINDEX search (locating 62 articles) by adding Criminal Justice Abstracts with full text as a second bibliographic database, it located 153articles.

5. *Use Boolean search logic.* It's often a good idea to narrow your search by requiring that abstracts contain combinations of words or phrases that include more of the specific details of your research question. Using the Boolean connector *AND* allows you to do this, whereas using the connector *OR* allows you to find abstracts containing different words that mean the same thing (see Figure 2.5).

FIGURE 2.5 ■ Starting a Search in Sociological Abstracts



6. *Use appropriate subject descriptors.* Once you have found an article you consider appropriate, check the "descriptors" or "subject terms" field in the citation. You can then redo your search after requiring that the articles be classified with some or all of these descriptor terms. It can be helpful to add "key terms" that have already been

attached to related articles by professional indexers (Gough et al. 2017:108–110). Such "controlled vocabularies" appear as "subject terms" in SocINDEX, as "thesaurus" terms in Sociological Abstracts, and as MeSH (medical subject headings) in MEDLINE. A librarian can help you to refine your search strategy in this way.

7. *Check the results.* Read the titles and abstracts you have found and identify the articles that appear to be most relevant. If possible, click on these article titles and generate a list of their references. See if you find more articles that are relevant to your research question but that you have missed so far. You will be surprised (I always am) at how many important articles your initial online search missed.

8. *Locate the articles.* Whatever database you use, the next step after identifying your references is to obtain the articles themselves. You will probably find the full text of many articles available online, but this will be determined by what journals your library subscribes to and the period for which it pays for online access. The most recent issues of some journals may not be available online. Keep in mind, your library will not have anywhere near all the journals (and books) that you run across in your literature search, so you will have to add another step to your search: checking the "holdings" information.

   If an article that appears to be important for your topic isn't available from your own library or online, you may be able to request a copy online through your library site or by asking a member of the library staff. You can also check http://worldcat.org to see what other libraries have the journal.

9. *Take notes on each article you read, organizing your notes into standard sections: theory, methods, findings, conclusions.* In any case, write your review of the literature so it contributes to your study in some concrete way; don't feel compelled to discuss an article just because you have read it. Be judicious. You are conducting only one study of one issue, and it will only obscure the value of your study if you try to relate it to every tangential point in related research.

Don't think of searching the literature as a one-time-only venture—something that you leave behind as you move on to your *real* research. You may encounter new questions or unanticipated problems as you conduct your research or as you burrow deeper into the literature. Searching the literature again to determine what others have found in response to these questions or what steps they have taken to resolve these problems can yield substantial improvements in your own research. There is so much literature on so many topics that it often is not possible to figure out in advance every subject for which you should search the literature or what type of search will be most beneficial.

Another reason to make searching the literature an ongoing project is that the literature is always growing. During the course of one research study, whether it takes only one semester or several years, new findings will be published and relevant questions will be debated. Staying attuned to the literature and checking it at least when you are writing up your findings may save your study from being outdated as soon as it is finished.

Searching the Web

The World Wide Web provides access to vast amounts of information of many different sorts (Ó Dochartaigh 2002). There are more than 1.8 billion websites, although only about one-quarter

(200,000,000) of these are active (http://www.internetlivestats.com/total-number-of-websites/). You can search the holdings of other libraries and download the complete text of government reports, some conference papers, and newspaper articles. You can find policies of local governments, descriptions of individual social scientists and particular research projects, and postings of advocacy groups. It's also hard to avoid finding a lot of information in which you have no interest, such as commercial advertisements, third-grade homework assignments, or advertisements galore.

So *caveat emptor* (buyer beware) is the watchword when you search the web. After all, it is a medium in which anyone with basic skills can post almost anything. On the other hand, the web provides access to resources of vital importance for many research projects; the trick is to use it well and wisely. Google has become the leading search engine. Its coverage is relatively comprehensive, and it does a good job of ranking search results by their relevancy (based on the terms in your search request). Google also allows you to focus your search just on images, discussions, or directories. There are several alternative search engines to Google—each has some advantages and disadvantages (Gil 2017), but at this time, most search needs can be met with Google.

Google Scholar is of special interest since it provides a publicly accessible tool for searching the scholarly literature across disciplines, but it also includes technical reports, theses, books, and other types of documents. Google Scholar found 26,700 documents in a search for "racial disparities in police shootings" [not in quotes] (on April 9, 2022), and since it lists articles in order of use of the search terms, frequency of citation, and other reasonable factors, the first several pages of citations provide a good way to identify potentially important omissions from your literature searches in bibliographic databases available at your library. However, in most cases, when you attempt to open the PDF of most of the full articles, you will be asked to pay for access. So you will still need to go through your library to obtain the full text of the articles that interest you (if your library subscribes to the source journals).

A *Google Scholar* search also produces a list of suggestions for related searches that can help you refine or expand your search. For example, the list of "related searches" at the bottom of the first page of my "racial disparities in police shootings" (without quotation marks) search includes "police shootings racial residential segregation." Clicking this link produced a list of 3,590 results.

On the web, less is usually more. As a general strategy, limit your inspection of websites to the first few pages that turn up in your list (they're ranked by relevance). See what those first pages contain, and then try to narrow your search by including some additional terms. If you limit your search to an exact phrase by putting the words in quotation marks, you can reduce the number of results dramatically and they are all likely to be related to exactly what you are looking for. Searching for "racial disparities in police shootings" [in quotes] produced only 81 results (compared to the 26,700 without quotes). But of course, you'll miss many related articles that way. You can limit the search more by limiting it to sites with the desired phrase in the title. For instance, the search *allintitle: racial disparities in police shootings* resulted in a dramatically smaller yield of results (6 in this case). You can also narrow your search by date (options in the left-hand margin). You'll find other ways of limiting your search in the "Advanced search" option in the *Google Scholar* menu.

It's all rather overwhelming, so you'll have to spend some time refining and trying out different sets of search terms. Before you begin, be sure to clarify the goals of your search. Will you check on the coverage of your literature searching? Review related government programs? Find reports and statistics about the research question? Examine commentary about it?

REVIEWING PRIOR RESEARCH

Once you have identified relevant information from prior research, it's time to review it. Focus your review on the articles you have found in peer-reviewed academic journals, so you know you are reading work by social scientists who subjected their reports to a quality review by other social scientists. It's not possible to overemphasize the importance of this step as you begin to investigate a social research question, so I will present the process in detail.

Consider some of the issues you can address if you use a literature review to refine your research focus:

- Has a consensus already emerged in response to this research question? If not, what are the differences in findings, and what remains to be investigated?
- What has not been acknowledged in prior publications? Have methods been fully described?
- What theories have been applied in this research, and which seem most consistent with the evidence?
- Which methods have been used in prior investigations? Have any difficulties emerged in their application?

If you keep issues like these in mind as you review the literature, you will create a solid foundation for your own research.

Although it's most important when you're starting out, reviewing the literature is also important at later stages of the research process. Throughout a research project, you will uncover new issues and encounter unexpected problems; at each of these times, you should search the literature to locate prior research on these issues and to learn how others responded to similar problems. Published research that you ignored when you were seeking other research on police shootings might become very relevant when you must decide which questions to ask people about their attitudes toward police and other authorities.

Your literature review will suggest specific research questions for further investigation and research methods with which to study those questions. Braga et al. (2018) learned from their literature review that prior research indicated—with just a few exceptions—that BWCs (body-worn cameras) reduce citizen complaints against officers. However, results of prior research on the impact of BWCs on officer use of force during police–citizen encounters had been inconsistent: some found a reduction in officer use of force and others found no change. One careful evaluation suggested the inconsistency might be due to differences in the extent to which officers complied with their department's BWC activation policy. In sites where officers notified citizens of BWC at the start of an encounter, use of force declined; in sites where they didn't make such an announcement, use of force increased.

Their literature review also found that prior research had not taken account of possible "spillover effects": the tendency of officers not wearing body cameras to change their behavior when working with other officers who were using BWCs. There was thus potential value in conducting new research that took account of both officer compliance and tested for spillover effects. In this way, reviewing the literature identifies unanswered questions and contradictory evidence.

Effective review of the prior research is thus an essential step in building the foundation for new research. You must assess carefully the quality of each research study, consider the implications of each article for your own plans, and expand your thinking about your research question

to account for new perspectives and alternative arguments. Through reviewing the literature and using it to extend and sharpen your own ideas and methods, you become a part of the social science community. Instead of being just one individual studying an issue that interests you, you are building on an ever-growing body of knowledge that is being constructed by the community of scholars.

Most of the research articles you find will include a short literature review on the specific focus of the research. Sometimes you'll find that someone else has already searched the literature on your research question and discussed what they found in a special review article or book chapter. All of these reviews can help, but they are no substitute for searching the literature yourself, selecting the articles and other sources that are most pertinent to your research question, and then reading what you have found. You are the only one who can decide what is relevant for your research question and the research circumstances you will be facing—the setting you will study, the timing of your study, the new issues you want to include in your study, and your specific methods. And you can't depend on any published research review for information on the most recent works. New research results about many questions appear continually in scholarly journals and books, in research reports from government agencies and other organizations, and on websites all over the world; you'll need to check for new research like this yourself.

The published scholarly journal literature can be identified in databases such as Sociological Abstracts, SocINDEX, PsycINFO, and MEDLINE. Because these literature databases follow a standard format and use a careful process to decide what literature to include, they are the sources on which you should focus. This section concentrates on the procedures you should use for reviewing the articles you find in a search of the scholarly literature, but these procedures can also be applied to reviews of research monographs—books that provide more information from a research project than can be contained in a journal article.

Reviewing the literature is really a two-stage process. In the first stage, you must assess each article separately. This assessment should follow a standard format such as that represented by the "Questions to Ask About a Research Article" in Appendix A. However, keep in mind that you can't adequately understand a research study if you just treat it as a series of discrete steps involving a marriage of convenience among separate techniques. Any research project is an integrated whole, so you must be concerned with how each component of the research design influenced the others—for example, how the measurement approach might have affected the causal validity of the researcher's conclusions and how the sampling strategy might have altered the quality of measures.

The second stage of the review process is to assess the implications of the entire set of articles (and other materials) for the different aspects of your research question and then to write an integrated narrative review that highlights these implications. Although you can find literature reviews that consist simply of assessments of one published article after another— that never get beyond the first stage in the review process—your understanding of the literature and the quality of your own work will be much improved if you develop an integrated review.

Reviewing the literature can itself become the research project! A "systematic review" begins with a very carefully designed search strategy and criteria for including research articles and then summarizes each included study with a detailed review protocol. A systematic review is published as a separate article rather than as part of the background in an article presenting the results of new research. Another way to evaluate systematically the findings of prior research is called "meta-analysis." In a meta-analysis, a researcher systematically identifies all prior quantitative research articles focused on a specific research question and then analyzes statistically

the combined findings from those prior studies. I will review the method of meta-analysis in Chapter 11.

In the next section, I show how you might answer many of the questions in Appendix A as I review a research article about the effects of aggressive policing. I will then show how the review of a single article can be used within an integrated review of the body of prior research relevant to a research question. I will conclude the material on literature reviews with a section on conducting a systematic review.

Single-Article Reviews: Aggressive Policing and Educational Performance of Minority Youth

Joscha Legewie at Harvard University and Jeffrey Fagan at Columbia Law School teamed up to study the effect of aggressive policing on the educational performance of New York City public school students in kindergarten to eighth grade. In this section, we will examine the article that resulted from that replication, which was published in the *American Sociological Review* (Legewie and Fagan 2019). The numbers in square brackets refer to the article review questions in Appendix A. Be sure to read each article review question in Appendix A when you read my answer to it.

The research question. Legewie and Fagan (2019:221) state their primary research question as the following: "What are the consequences of the increasing presence of police in minority communities for minority youths' educational performance?" They focus specifically on the effect of aggressive, order-maintenance policing (also known as "broken-windows policing") on African American boys and in comparison to Hispanic boys and girls in both groups. They also propose to examine the mechanisms by which any effects occur. [1]. The purpose of the study was explanatory: Its goal was to explain variation in educational performance[2]. Theoretical models of neighborhood effects and policing guide the study, pointing to different mechanisms that could shape the effects of policing on educational performance. Their theorizing does not make clear connections to such potentially relevant theories as deterrence theory, labeling theory, or conflict theory. The research is applied and they approach it with a positivist philosophy that focuses on cause and effect connections. The research is not guided by their values, although their conclusions suggest they favor policies that would improve rather than diminish educational performance [3].

Legewie and Fagan review a large body of prior research literature, giving particular attention to research that has tested different ways in which policing practices could affect students' educational performance [4]. Figure 2.5 (preceding) shows what Legewie and Fagan might have entered on their computer if they had searched Sociological Abstracts to find research related to aggressive policing and educational performance.

The research uses only administrative records, so there is no direct contact with human subjects and no discussion of ethical guidelines [6].

The research design. The explanatory focus of the research identifies it as deductive [5]. The primary study hypothesis is that aggressive policing strategies and tactics lower educational performance and perpetuate racial inequalities in educational outcomes, although it is not labeled explicitly as a hypothesis. Based on extensive prior research, three other hypotheses are proposed as mechanisms that might explain the effect of aggressive policing on educational performance: (a) Aggressive policing reduces crime rates, which in turn may improve educational performance; (b) aggressive policing may reduce trust in institutions and school attendance, which may negatively influence educational performance; (c) aggressive policing may trigger adverse health effects, such as stress and fear, which reduce educational performance [7].

The independent variable in the primary hypothesis is aggressive policing and the dependent variable is educational performance. The secondary hypotheses are really each a pair of hypotheses; in the first of each pair, aggressive policing is initially the independent variable and crime rate, trust, and poor health are the dependent variables, while in the second pair, these three "dependent" variables become independent variables that predict change in educational outcomes—the dependent variable. Hypotheses are also proposed involving differences in effects in relation to student age, race, and gender.

The major concepts in the study, aggressive policing and educational performance, were defined clearly [9] and then measured with straightforward indicators—days of exposure of a school neighborhood to Operation Impact (New York City police department's aggressive policing program) as the measure of aggressive policing and standardized test scores (English and math) as the measure of educational performance. Previously collected data are used to measure neighborhood crime level, positive school attitudes and trust, and school attendance. However, there are no indicators of poor health, such as stress levels or fear, in spite of an extensive section of the literature review devoted to these issues [9]. The data were obtained from police department and school administrative records, including results of a student survey. These data sources are described in detail, but there is no indication of efforts to establish their reliability or validity [10].

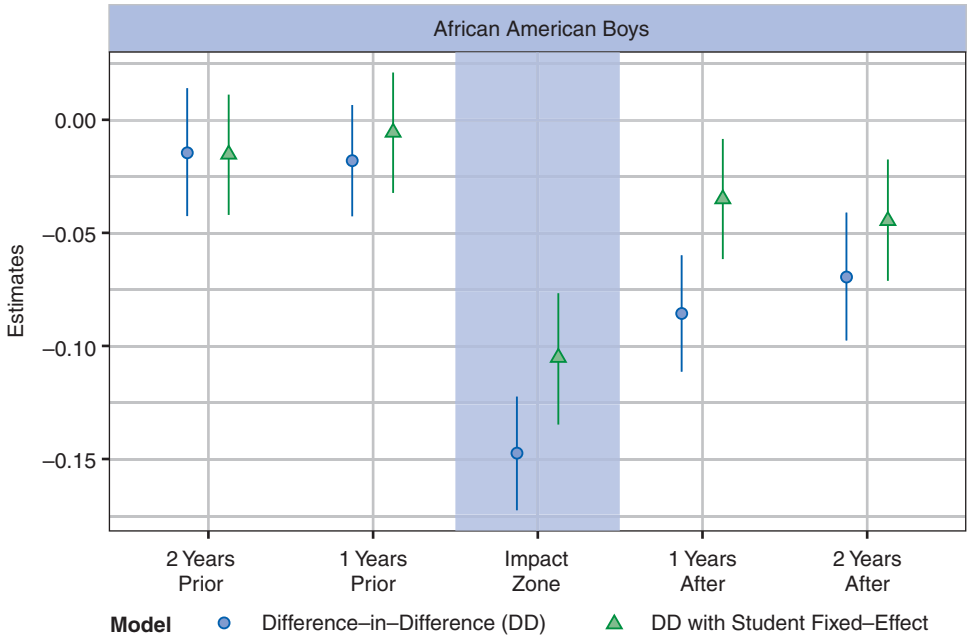
The sample consisted of all 285,439 African American and Hispanic students attending schools in zones designated as Operation Impact zones at some time between 2003 and 2012. It was thus really the entire population of these students rather than a sample from the population. Too few white and Asian students lived in these areas to be able to analyze them separately, so these students are excluded from the analysis. The analysis was also limited to students between 9 and 15 years old, for the same reason. A small fraction of students who did not participate in the yearly state test of educational performance or for whom data were missing on other variables were also excluded from the analysis [11]. The authors report that 2.6% of the cases had data missing on one or more variables [12].

Individual students were the units of analysis, although most of the predictors are based on school- or neighborhood-level data. There is some risk of a type of ecological fallacy since there is no individual-level measure of exposure to Operation Impact—but this threat is not discussed [13]. The study used a longitudinal panel design since students were followed over time, before and after the period when police used Operation Impact [14]. Use of a type of quasi-experimental design (before–after) strengthened the test of the authors' causal hypotheses. In the article's limitations section, the authors consider whether the findings would hold up in other cities, suburban or rural areas, and across the world. They recognize that more research would be needed to determine such external generalizability [15].

The research findings and conclusion. Legewie and Fagan (2019) analyzed only quantitative data, so they did not use mixed methods [18]. The project relied on secondary data from the New York City Department of Education (NYCDOE) and the New York Police Department (NYPD). The authors do not discuss the quality of these data, apparently assuming that such administrative data collected in these large, professionally run organizations is accurate. Use of this "big data" allows sophisticated analyses of changes over time. No information is provided on protection of anonymity of individual records, but it is likely that such steps were taken [19]. The design was not historical-comparative and did not involve a content analysis [20]. The authors highlighted the importance of social context throughout the article and gave careful attention to subjective meanings by using survey data that indicated students' level of trust toward schools and teachers [21]

The central finding is that the primary hypothesis is supported: launching Operation Impact (aggressive policing) in an area was followed by a decline in test scores of African American boys in public schools in that area; an apparent effect that diminished in size over the following two years (see Figure 2.6). This is certainly an important finding. Legewie and Fagan then present tests of their secondary hypotheses and find that the effect of aggressive policing was not due to an increase in the crime rate in these areas; however, a decline in school attendance for African American boys during the Operation Impact period seemed to be partly responsible for the decline in test scores.

FIGURE 2.6 ■ Effect of Aggressive Policing (Operation Impact) on English Test Scores Before, During, and After Operation Impact



Source: Legewie, Joscha, and Jeffrey Fagan. 2019. "Aggressive Policing and the Educational Performance of Minority Youth." *American Sociological Review* 84(2):220–47.

The statistical analyses required to distinguish effects of changes in the crime rate from the effect of Operation Impact are complicated but presented clearly [22]. The authors represent the findings well in the conclusions and carefully consider alternative interpretations [23]. Overall, the study is a major advance over prior research [24] and suggests many additional questions for further research [25], such as: Would implementing a community policing strategy have the same type of effects on school performance as the use of aggressive policing? Do youth with stronger ties to family and/or community groups suffer less in their academic performance after exposure to aggressive policing?

Integrated Literature Reviews: When Does Arrest Matter?

The goal of the second stage of the literature review process is to integrate the results of your separate article reviews and develop an overall assessment of the implications of prior research. The integrated literature review should accomplish three goals: (1) summarize prior research,

(2) critique prior research, and (3) present pertinent conclusions (Hart 1998:186–187). I'll discuss each of these goals in turn.

1. *Summarize prior research.* Your summary of prior research must focus on the particular research questions you will address, but you may also need to provide some more general background. Carolyn Hoyle and Andrew Sanders (2000:14) begin their *British Journal of Criminology* research article about mandatory arrest policies in domestic violence cases with what they term a "provocative" question: What is the point of making it a crime for men to assault their female partners and ex-partners? The researchers then review the different theories and supporting research that have justified different police policies: the "victim choice" position, the "pro-arrest" position, and the "victim empowerment" position. Finally, Hoyle and Sanders review the research on the "controlling behaviors" of men that frames the specific research question on which they focus: how victims view the value of criminal justice interventions in their own cases (Hoyle and Sanders 2000:15).

   After you have drafted your summary of the literature, ask yourself three questions about it:

   a. Have you been selective? If there have been more than a few prior investigations of your research question, you will need to narrow your focus to the most relevant and highest quality studies. Don't cite a large number of prior articles "just because they are there."
   b. Is the research up-to-date? Be sure to include the most recent research, not just the "classic" studies.
   c. Have you used direct quotes sparingly? To focus your literature review, you need to express the key points from prior research in your own words. Use direct quotes only when they are essential for making an important point (Pyrczak 2005:51–59).

2. *Critique prior research.* Evaluate the strengths and weaknesses of the prior research. In addition to all the points that you develop as you answer the article review questions in Appendix A, you should also select articles for review that reflect work published in peer-reviewed journals and written by credible authors who have been funded by reputable sources. Consider the following questions as you decide how much weight to give each article:

   a. How was the report reviewed before its publication or release? Articles published in academic journals go through a rigorous review process, usually involving careful criticism and revision. Top refereed journals may accept only 10% of the submitted articles, so they can be very selective. Dissertations go through a lengthy process of criticism and revision by a few members of the dissertation writer's home institution. A report released directly by a research organization is likely to have had only a limited review, although some research organizations maintain a rigorous internal review process (see discussion of research organizations, later in this chapter). Papers presented at professional meetings may have had little prior review. Needless to say, more confidence can be placed in research results that have been subject to a more rigorous review.
   b. What is the author's reputation? Reports by an author or team of authors who have published other work on the research question should have somewhat greater credibility at the outset.

   c. Who funded and sponsored the research? Major federal funding agencies and private foundations fund only research proposals that have been evaluated carefully and ranked highly by a panel of experts. These agencies also often monitor closely the progress of the research. This does not guarantee that every such project report is good, but it goes a long way toward ensuring some worthwhile products. Conversely, research that is funded by organizations that have a preference for a particular outcome should be given particularly close scrutiny (Locke, Silverman, and Spirduso 1998:37–44).

3. *Present pertinent conclusions.* Don't leave the reader guessing about the implications of the prior research for your own investigation. Present the conclusions you draw from the research you have reviewed. As you do so, follow several simple guidelines:
   a. Distinguish clearly your own opinion of prior research from the conclusions of the authors of the articles you have reviewed.
   b. Make it clear when your own approach is based on the theoretical framework that you use and not on the results of prior research.
   c. Acknowledge the potential limitations of any empirical research project (Pyrczak 2005:53–56).
   d. Explain how the unanswered questions raised by prior research or the limitations of methods used in prior research make it important for you to conduct your own investigation (Fink 2005:190–192).

A good example of how to conclude an integrated literature review is provided by an article based on a study in Milwaukee of the police response to domestic violence cases. For this article, Ray Paternoster and his colleagues (1997) sought to determine whether police officers' use of fair procedures when arresting assault suspects would lessen the rate of subsequent domestic violence. Paternoster et al. (1997) conclude that there has been a major gap in the prior literature: "Even at the end of some seven experiments and millions of dollars, then, there is a great deal of ambiguity surrounding the question of how arrest impacts future spouse assault" (p. 164). Specifically, the researchers note that each of the seven experiments focused on the effect of arrest itself but ignored the possibility that "particular kinds of police procedure might inhibit the recurrence of spouse assault" (p. 165).

So Paternoster and his colleagues (1997) ground their new analysis in additional literature on procedural justice and conclude that their new analysis will be "the first study to examine the effect of fairness judgments regarding a punitive criminal sanction (arrest) on serious criminal behavior (assaulting one's partner)" (p. 172).

Systematic Literature Reviews: Second Responder Programs and Repeat Family Abuse Incidents

Any literature review should be systematic, but the term **systematic review** designates an approach to literature review that is much more structured than what is usually done in an integrated review of the literature. It is best thought of as a distinct method of research, for it often is itself the basis of a published article—rather than being used to provide the background in an article that reports the results of a new research study. Professional librarians often play a major role in systematic reviews and may coauthor the resulting article. There are also helpful tools available at https://systematicreviewsjournal.biomedcentral.com/. Published systematic reviews

are now archived on searchable websites, including http://www.cochranelibrary.com/ (health care), https:// https://www.campbellcollaboration.org/better-evidence (social interventions), and http://eppi.ioe.ac.uk/cms/ (multiple topic areas).

Developing a systematic review involves all of the activities I have just described for searching and reviewing the literature, but it proceeds with more explicit plans and reports at each stage of the process. Many systematic review efforts adhere to a set of rules known as the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses): http://www.prisma-statement.org/Default.aspx.

Stephen Puntis and his colleagues (2018) in the United Kingdom conducted a systematic review of the effect of adding mental health interventions, called "street triage interventions," to police responses to incidents with a suspected mental health component. The article reporting the results illustrates each stage of a systematic review.

1. *Define one or more specific research questions.* Puntis et al. (2018) specify three aims:
   a. To identify and describe different *co*-response models of police mental health street triage
   b. To identify demographic and clinical characteristics of service users
   c. To evaluate the evidence for the effectiveness of co-response police mental health triage

2. *List the terms to be used in searching, the specific sources to be searched, and the inclusion criteria.* After a detailed description of the criteria for inclusion of studies, Patterson and Swan identified the following 9 bibliographic databases to search:
   a. Ovid
   b. MEDLINE
   c. Embase
   d. PsycINFO
   e. EBSCO CINAHL
   f. Scopus
   g. Thompson Reuters Web of Science Core Collection
   h. The Cochrane Library
   i. ProQuest National Criminal Justice Reference Service Abstracts

      They also searched the "grey literature": meaning reports of research that have not been published in peer-reviewed articles.
   j. ProQuest Dissertations & Theses
   k. EThoS
   l. OpenGrey

      The following keywords were used to search the databases listed above:
   a. street triag*
   b. policing or police and triag*
   c. "mental health" and triag*
   d. "liaison and diversion"
   e. "specialized mental health response"
   f. "co-responder"

      and a few others.

TABLE 2.2 ■ Systematic Review Inclusion and Exclusion Criteria

Criteria	Sub-criteria	
	Inclusion	**Exclusion**
1. A co-response model of police mental health triage	• Include both police officers and mental health workers in the response • Be a response to a police incident	
2. Describe a triage model or model development OR implementation of a triage model OR epidemiological study OR evaluation of effectiveness	• Any original published article • Grey literature, service evaluations	• Review articles, book chapters, editorials, or comments
3. English language article		

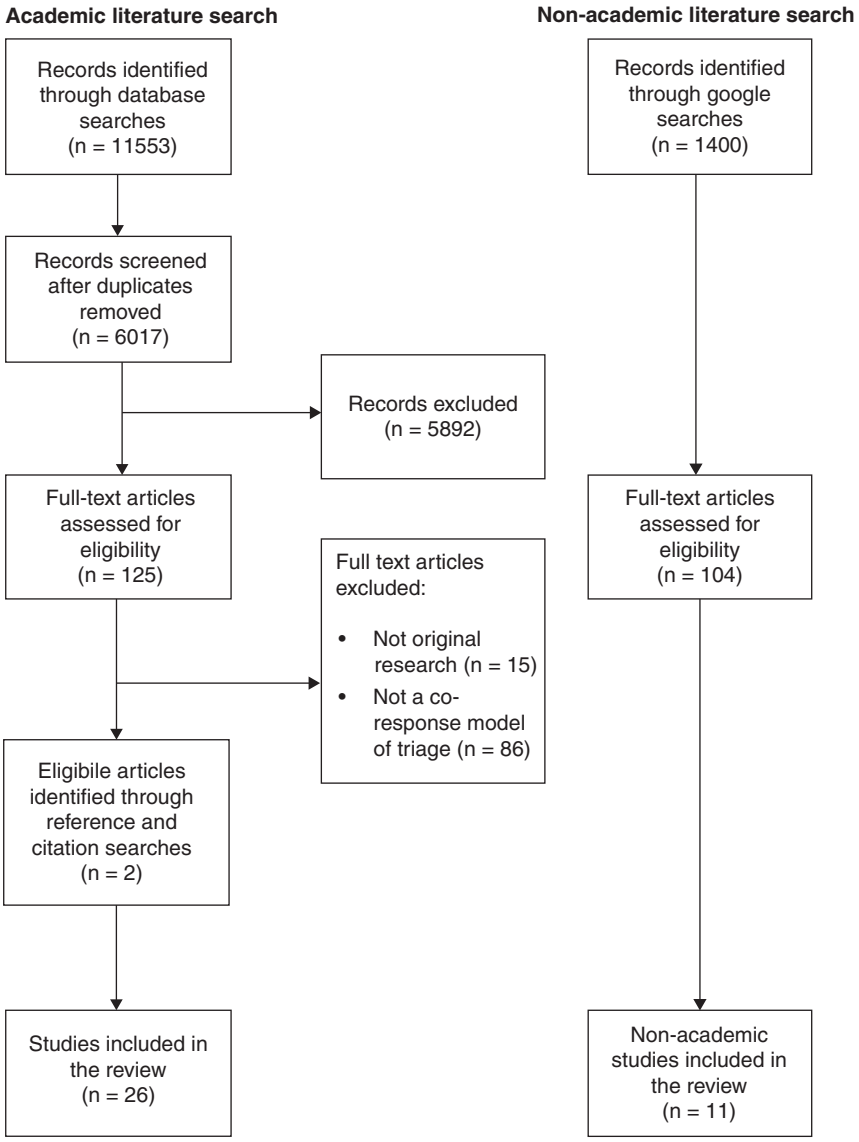
[the * means that the word it is attached to, like "shoot" could include different endings, such "er" or "ing"]

Puntil et al. (2018) established detailed inclusion criteria (see Table 2.2):

3. *Report the results of the search and the selection process*, often in the type of flow diagram recommended by PRISMA (see Figure 2.7). Puntis et al. initially identified 11,553 published articles using these search terms, but after removing duplicates and screening title and abstracts for relevance, they concluded that only 125 warranted full-text review. These decisions were made after review by two teams of reviewers operating independently and then comparing results before discussion to resolve discrepancies with a third reviewer. Of the articles reviewed in this way, just 26 met the detailed criteria for inclusion. Another 11 relevant studies were identified in the grey ("non-academic") literature (see Figure 2.7).
4. *Code the characteristics of the selected studies.* A broad range of characteristics were coded, including age and race of the person involved in the incident, reason for the triage response, and reduction in police detentions (see Table 2.3 for some of the coded characteristics).
5. *Summarize the results in a narrative review.* Puntis et al. (2018) conclude "that co-response models of mental health police triage may reduce the use of police powers of detention" with persons with mental illness but highlight a lack of adequate quality in prior research. They identify unanswered questions, such as whether a police response is most appropriate in these cases.

Systematic reviews provide an excellent source of information about prior research in areas in which interventions have been tested and such reviews have been conducted. Be sure to read any that are available pertaining to your research question.

FIGURE 2.7 ■ PRISMA Flow Diagram


Academic literature search
Non-academic literature search
Records identified through database searches (n = 11553)
Records identified through google searches (n = 1400)
Records screened after duplicates removed (n = 6017)
Records excluded (n = 5892)
Full-text articles assessed for eligibility (n = 125)
Full text articles excluded:
• Not original research (n = 15)
• Not a co-response model of triage (n = 86)
Full-text articles assessed for eligibility (n = 104)
Eligibile articles identified through reference and citation searches (n = 2)
Studies included in the review (n = 26)
Non-academic studies included in the review (n = 11)

Source: Puntis, Stephen, Devon Perfect, Abirami Kirubarajan, Sorcha Bolton, Fay Davies, Aimee Hayes, Eli Harriss, and Andrew Molodynski. 2018. "A Systematic Review of Co-Responder Models of Police Mental Health 'street' Triage." *BMC Psychiatry* 18(1):1–11.

TABLE 2.3 ■ Suggested Framework for Collecting and Reporting Co-response Triage Models [partial table]

Topic	Variable to Report	Description of Variable and Reason for Reporting
Identify as co-response triage service	Name of service	Give the name of the service to enable grouping of models and their comparison.
	Identify as co-response	Include a sentence in the service description to signpost to readers and researchers that the service is a co-response model of police mental health triage.
Model characteristics	Model of co-response	Define the model of how the mental health professional assists the police officer during the incident. *Ride-along* models are those in which the police officer (PO) and mental health (MH) worker attend the incident together in a vehicle. *Ride-separate* models are when PO and MH worker arrive at the incident separately. *Telephone support* models are those in which the MH worker provides assistance via a telephone or radio. Services that provide a combination of these models should specify when each model is used.
	Method of referral to co-response team	Describe how crisis incidents are referred to the triage team (e.g., emergency response, direct from police officers, publicly available direct phone line), and from whom triage can take referrals (e.g., police officers, other emergency services, mental health services, the public, etc.).

Source: Puntis, Stephen, Devon Perfect, Abirami Kirubarajan, Sorcha Bolton, Fay Davies, Aimee Hayes, Eli Harriss, and Andrew Molodynski. 2018. "A Systematic Review of Co-Responder Models of Police Mental Health 'street' Triage." *BMC Psychiatry* 18(1):1–11.

RESEARCH IN THE NEWS
FATAL FORCE: A PUBLIC DATABASE OF POLICE SHOOTINGS

After the protests and nationwide soul-searching about police shootings of Black Americans that began with Michael Brown's 2014 death in Ferguson, Missouri, it is hard to believe these events were not often included in official records. Yet *The Washington Post* found that many police departments had not been reporting these events to the FBI for inclusion in the official record, so *The Washington Post* began a database of fatal police shootings in 2015. Compiling records from police, news, and social media reports, they have found about 1,000 fatal police shootings each year since 2015. Black Americans were killed in these shootings more than twice as often relative to their share of the population than white Americans and Hispanic persons were killed slightly less than twice as often. More than 95% were men, and more than half were between 20 and 40 years old.

For Further Thought?

1. Do you believe police department record-keeping practices should be improved to ensure an official record of these incidents? What opportunities and obstacles might be encountered in such an effort?
2. What type of research could improve our understanding of the differences in police shootings by race and ethnicity?

News source: Tate, Julie, Jennifer Jenkins, and Steven Rich. 2021. "Fatal Force." *The Washington Post*, August 30. https://www.washingtonpost.com/graphics/investigations/police-shootings-database/.

SOCIAL RESEARCH STRATEGIES

With a research question formulated, a review of the pertinent literature taking shape, and a theoretical framework in mind, you are ready to consider the process of conducting your research.

When you conduct social research, you are attempting to connect theory with empirical data—the evidence obtained from the social world. Researchers may make this connection by starting with a social theory and then testing some of its implications with data. This is the process of deductive research; it is most often the strategy used in quantitative methods. Alternatively, researchers may develop a connection between social theory and data by first collecting the data and then developing a theory that explains the patterns in the data (see Figure 2.8). This inductive research process is more often the strategy used in qualitative methods. As you'll see, a research project can draw on both deductive and inductive strategies.

Social theories do not provide the answers to the questions we pose as topics for research. Instead, social theories suggest the areas on which we should focus and the propositions that we should consider for a test. Figure 2.9 summarizes how the two theories that guided Braga et al.'s (2018) research and the theory that guided Paternoster et al.'s (1997) analysis relate to the question of whether body-worn cameras change police behavior. By helping us make such connections, social theory sensitizes us to the possibilities, thus helping us design better research and draw out the implications of our results. Before, during, and after a research investigation, we need to keep thinking theoretically.

FIGURE 2.8 ■ The Links Between Theory and Data

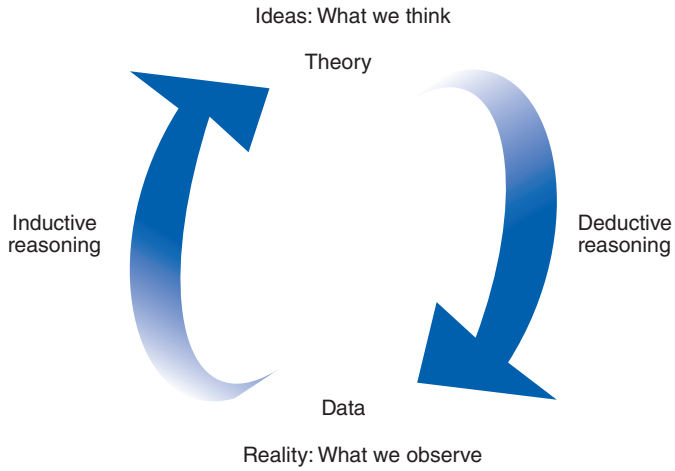


FIGURE 2.9 ■ Two Social Theories Predicting Body-Worn Cameras Change Police Behavior

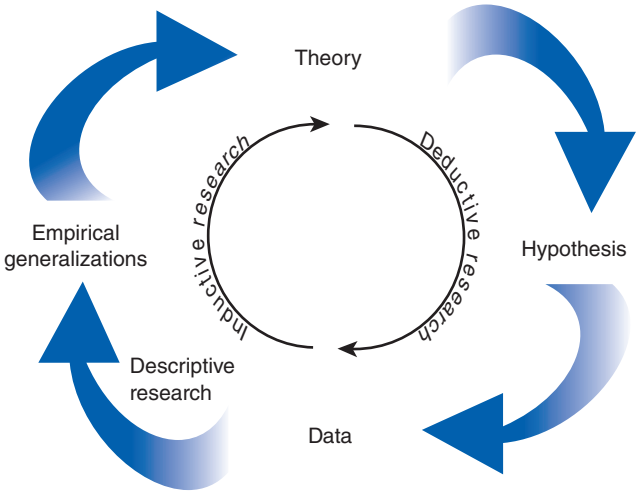
	Rational choice theory	Self-awareness theory
Theoretical assumption	People's behavior is shaped by calculations of the costs and benefits of their actions.	Self-aware people more likely to align behavior with personal standards
Criminological component	Deterrence theory: People break the law if the benefits of doing so outweigh the costs.	When under observation, people exhibit more socially acceptable behavior
Prediction (effect of arrest for domestic assault)	BWCs increase likelihood of capturing on video any misconduct, thereby increasing perceived likelihood of apprehension and rapidity of punishment.	BWCs during police-citizen interaction increases socially desirable behaviors.

Source: Image based on text in Braga, Anthony A., William H. Sousa, James R. Coldren, and Denise Rodriguez. 2018. "The Effects of Body-Worn Cameras on Police Activity and Police-Citizen Encounters: A Randomized Controlled Trial." *Journal of Criminal Law and Criminology* 108(3):515–516.

Explanatory Research

The process of conducting research designed to test explanations for social phenomena involves moving from theory to data and then back to theory. This process can be characterized with a **research circle** (Figure 2.10).

FIGURE 2.10 ■ The Research Circle



Deductive Research

As Figure 2.10 shows, in **deductive research**, a specific expectation is deduced from a general theoretical premise and then tested with data that have been collected for this purpose. We call the specific expectation deduced from the more general theory a **hypothesis**. It is the hypothesis that researchers actually test, not the complete theory itself. A hypothesis proposes a relationship between two or more **variables**—characteristics or properties that can vary.

Example of a hypothesis: The higher the poverty rate in a community, the higher the percentage of community residents who are homeless.

Example of a variable: The degree of honesty in verbal statements.

Variation in one variable is proposed to predict, influence, or cause variation in the other. The proposed influence is the **independent variable**; its effect or consequence is the **dependent variable**. After the researchers formulate one or more hypotheses and develop research procedures, they collect data with which to test the hypothesis.

Example of an independent variable: Poverty rate.

Example of a dependent variable: Percentage of community residents who are homeless.

Hypotheses can be worded in several different ways, so identifying the independent and dependent variables is sometimes difficult. When in doubt, try to rephrase the hypothesis as an *if–then* statement: "*If* the independent variable increases (or decreases), *then* the dependent variable increases (or decreases)." Table 2.4 presents several hypotheses with their independent and dependent variables and their if–then equivalents.

TABLE 2.4 ■ Examples of Hypotheses

Original Hypothesis	Independent Variable	Dependent Variable	If–Then Hypothesis	Direction of Association
1. The greater the use of the Internet, the greater the strength of distant family ties.	Level of Internet use	Strength of distant family ties	If Internet use is greater, then the strength of distant family ties is greater.	+
2. The risk of property theft decreases as income increases.	Income	Risk of property theft	If income is higher, then the risk of property theft is less.	–
3. If years of education decrease, income decreases.	Years of education	Income	If years of education decrease, then income decreases.	+
4. Political conservatism increases with income.	Income	Political conservatism	If income increases, then political conservatism increases.	+
5. Property crime is higher in urban areas than in suburban or rural areas.	Type of community	Rate of property crime	If areas are urban, then property crime is higher compared with crime in suburban or rural areas.	NA

Table 2.4 demonstrates another feature of hypotheses: **direction of association**. When researchers hypothesize that as values on one variable increase the values on the other variable increase, the direction of association is positive (Hypotheses 1 and 4). When one variable decreases as the other variable decreases, the direction of association is also positive (Hypothesis 3). But when values of one variable increase as the values of the other decrease or vice versa, the direction of association is negative, or inverse (Hypothesis 2). Hypothesis 5 is a special case, in which the independent variable is qualitative: It cannot be said to increase or decrease. In this case, the concept of direction of association does not apply, and the hypothesis simply states that one category of the independent variable is associated with higher values on the dependent variable.

Both explanatory and evaluative studies are types of deductive research. The study of effects of body-worn cameras by Braga et al. (2020) was an evaluative study because the researchers sought to explain how a government policy requiring police to wear cameras would affect police–citizen encounters and police work activities. The researchers deduced from deterrence theory the expectation that BWCs would deter crimes by both police and citizens

by increasing the likelihood of discovery and sanction. They then collected data to test this expectation.

In both explanatory and evaluative research, the statement of expectations for the findings and the design of the research to test these expectations strengthen the confidence we can place in the test. Deductive researchers show their hand or state their expectations in advance and then design a fair test of those expectations. Then, "the chips fall where they may"—in other words, the researcher accepts the resulting data as a reasonably objective picture of reality.

Body-Worn Cameras and the Research Circle

Concern with the misapplication of police powers, including the incidence of police shootings, has resulted in many calls for police officers to be required to use body-worn cameras. But does the use of BWCs improve officer behavior and reduce the likelihood of police shootings and less harmful forms of aggressive policing? I will illustrate how the research circle works with some social science research conducted to answer this question. I will begin with an experimental study in Las Vegas designed by criminologist Anthony A. Braga, William H. Sousa, James R. Coldren Jr., and Denise Rodrigues 2018).

Braga et al.'s (2018) study was designed to test a hypothesis. According to deterrence theory, punishment will reduce recidivism, or the propensity to commit further crimes. From this theory, Braga et al. (2018:516) deduced the specific hypothesis that a BWC will reduce the likelihood of misconduct. In this hypothesis, wearing a body camera is the independent variable, and the likelihood of misconduct is the dependent variable (it is hypothesized to depend on the BWC). They expected that BWCs will have this deterrent effect because they increase "perceptions of the likelihood of apprehension and celerity [swiftness] of punishment." Braga et al. also suggest that another theory, self-awareness theory, also predicts this effect of the use of BWCs.

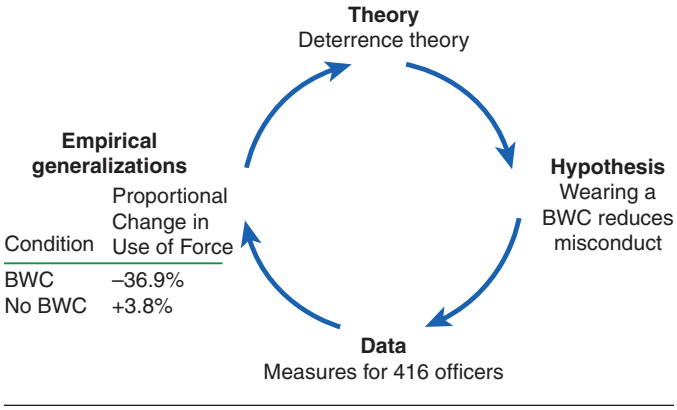
Of course, in another study, wearing body cameras might be the dependent variable in relation to some other independent variable. For example, in the hypothesis "The higher the rate of crime in a community, the lower the likelihood of police wearing body cameras," the independent variable is the crime rate and the dependent variable is the likelihood of wearing BWCs. Only within the context of a hypothesis, or a relationship between variables, does it make sense to refer to one variable as dependent and the other as independent.

Braga et al. (2018) designed a test of their hypothesis with funding from the U.S. Department of Justice and in collaboration with the Las Vegas Metropolitan Police Department (LVMPD). In their experiment, 416 police officer volunteers were randomly assigned to wear a body camera or not to wear one for twelve months. As you'll learn in Chapter 6, the random assignment ensured there were no systematic differences in the characteristics of the officers who wore body cameras and those who didn't. The LVMPD also provided the researchers with detailed information on the characteristics of the officers who did not participate in the experiment. For the one-year period after the experiment began, the researchers tracked with department records the incidents that officers responded to, citizen complaints filed against officers, and use of force reports filed by the officers.

When the researchers examined the full year of data (police records for the officers in their experiment), they found that among those assigned to the treatment (BWC) condition, the decline in citizen complaints was 25.0% larger than for the control (no-BWC) condition. The percentage of officers in the BWC group who filed at least one use of force report declined by 40.7% more than in the no-BWC group. These patterns in the data, or **empirical generalizations**, were consistent with the hypothesis that the researchers deduced from deterrence theory.

The theory thus received support from the experiment (see Figure 2.11). Braga et al. (2018) also found the officers who had volunteered to participate in the experiment did not differ in their characteristics from those who did not volunteer, thus making it likely the findings could be generalized to the population of police officers in the LVMPD.

FIGURE 2.11 ■ The Research Circle: Body-Worn Cameras



Source: Data from Braga et al. 2018:534.

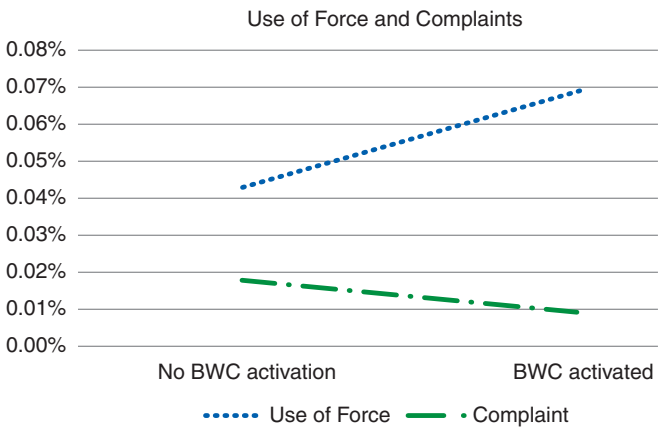
But there were many reasons for Braga and his coauthors to be concerned with the validity of the effects they found and the uniqueness of the conditions for their study. In another article, Sousa et al. (2016) identified major challenges they encountered in the NVMPD study: potential "contamination" when control group officers (non-BWC) joined treatment group officers (BWC) in the response to a call; attrition from the experiment when officers moved to other positions; unavailability of the required technological supports for BWCs in some areas; allowing officer participation in the experiment to be voluntary. They adjusted some of their study procedures to at least partially overcome some of these challenges but also encouraged other researchers to initiate new research that would build on their experiences.

Most of the researchers who studied the effects of BWCs after the Braga et al. (2018) study probably did not do so in response to the Sousa et al. (2016) article, but we can think of the later BWC studies as representing additional trips around the research circle, some of them intended to **replicate** (repeat) the Braga et al. (2018) experiment in different cities and others simply pushing forward with related research questions and/or enhanced research designs.

Jessica Huff, Katz, and Hedberg (2020) studied the effect of BWCs in Phoenix. Like Braga et al. (2018), they used an experimental design, but in addition to using officers who volunteered to be in the experiment, they also added a mandate for some officers to participate in a later phase of the experiment. Unlike Braga et al. (2018), they also measured whether officers activated their BWC during each incident, and they distinguished different types of incidents (such as violent offense, property offense, and vehicle stops).

When they analyzed their data, Huff et al. (2020) found that BWC activation made a difference: When the BWC was activated, use of force reports increased and the frequency of citizen complaints decreased (see Figure 2.12).

FIGURE 2.12 ■ Effect of BWC Activation on Use of Force and Complaints



In a new study in collaboration with the Boston Police Department, Anthony A. Braga and a different set of collaborators (2020) found that use of BWCs in a district generated "spillover effects" on police officers in those districts who were aware that BWCs might be present in an encounter. Another experimental study of BWCs in Australia, by Joseph Clare and his collaborators (2021), found that use of BWCs did not reduce the likelihood of citizen complaints or of filing of a use-of-force report and did not prove to have much impact on the outcomes of court cases resulting from the incidents. However, in this study, use of BWCs did increase the likelihood of police filing criminal charges and restraining orders against alleged offenders.

Thus, as studies about BWCs continued, the accumulating evidence indicated the use of BWCs often lessened the use of aggressive policing, but its effects varied with particular outcomes in relation to other aspects of the way the BWCs were deployed and with some differences apparently related to the location of use.

Inductive Research

In contrast to deductive research, **inductive research** begins with specific data, which are then used to develop (induce) a general explanation (a theory) to account for the data. One way to think of this process is in terms of the research circle: Rather than starting at the top of the circle with a theory, the inductive researcher starts at the bottom of the circle with data and then develops the theory. Another way to think of this process is represented in Figure 2.13. In deductive research, reasoning from specific premises results in a conclusion that a theory is supported, but in inductive research, the identification of similar empirical patterns results in a generalization about some social process.

Inductive reasoning enters into deductive research when we find unexpected patterns in the data we have collected for testing a hypothesis. We may call these patterns **anomalous findings**. When these unexpected patterns lead to new explanations, insights, or theoretical approaches, we call them **serendipitous findings**. However, the adequacy of an explanation formulated after the fact is necessarily less certain than an explanation presented before the collection of data and tested in a planned way with the data. Every phenomenon can always be explained in *some* way. Inductive explanations are thus more trustworthy if they are confirmed subsequently with deductive research.

FIGURE 2.13 ■ Deductive and Inductive Reasoning

Deductive

Premise 1:	All unemployed spouse abusers recidivate.
Premise 2:	Joe is an unemployed spouse abuser.
Conclusion:	***Joe will recidivate.***

Inductive

Evidence 1:	Joe, an unemployed spouse abuser, recidivated.
Evidence 2:	Harold, an unemployed spouse abuser, recidivated.
Evidence 3:	George, an employed spouse abuser, didn't recidivate.
Conclusion:	***All unemployed spouse abusers recidivate.***

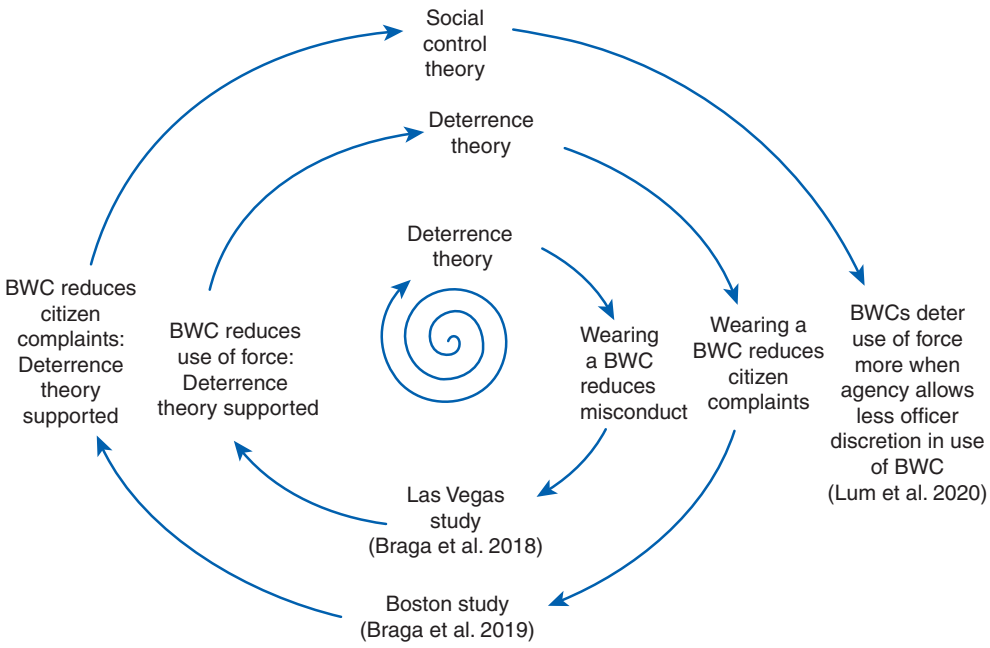
An Inductive Approach to Explaining Support for Black Lives Matter

Miller, O'Dea, and Donald (2021) hypothesized that higher levels of anti-Black sentiment and lower levels of belief that white privilege exists or that racism is a systemic problem would be associated with less support for Black Lives Matter. They tested these and related hypotheses with a sample of 225 participants recruited for online surveys. They measured their independent variables with multi-question indexes like the Modern Racism Scale (McConahay 1986) and the White Privilege Attitudes Scale (Pinterits, Paul, and Spanierma 2009). Measures of their dependent variables included three questions about the extent of support for Black Lives Matter or for two phrases that are used to indicate disagreement with the BLM movement: All Lives Matter and Blue Lives Matter. They tested their hypotheses and found them confirmed. All very deductive so far.

But their research took an inductive turn when Miller et al. (2021) wondered whether respondents' conceptualizations of prejudice as a systemic or individual issue would predict support for efforts to address social inequalities between white people and people of color. In this exploratory analysis, they found that when they compared people having similar levels of beliefs that racism is an individual problem, those with stronger beliefs that racism is a systemic problem tended to perceive more inequality in the United States, were more supportive of Black Lives Matter and less supportive of Blue Lives Matter and All Lives Matter, and were more supportive of white people and people of color protesting inequality. The researchers also explored differences in perceptions of how white people and people of color should protest inequality in the U.S. and found multiple patterns of influence on these protest beliefs. Although these associations were not hypothesized in advance of the analysis, they can enrich the interpretation of the hypothesized associations and suggest directions for future research.

The investigations conducted to answer the general research question about the impact of BWCs have generated a substantial body of scholarship that provides a picture of the generalizability of findings and that addresses many more specific questions about conditions that affect the impact of BWCs. Researchers have now traversed the research circle multiple times in these investigations, a process better described as a spiral (see Figure 2.14). Many investigations have traversed the research circle in a deductive, hypothesis-testing way. They started with theory and then deduced and tested hypotheses. Others have been more inductive or have had an inductive component: They started with empirical generalizations from the data they had already obtained and then turned to a new theory to account for the unexpected patterns in the data.

FIGURE 2.14 ■ The Research Spiral: Body-Worn Camera Experiment


Social control theory
Deterrence theory
Deterrence theory
BWC reduces citizen complaints: Deterrence theory supported
BWC reduces use of force: Deterrence theory supported
Wearing a BWC reduces misconduct
Wearing a BWC reduces citizen complaints
BWCs deter use of force more when agency allows less officer discretion in use of BWC (Lum et al. 2020)
Las Vegas study (Braga et al. 2018)
Boston study (Braga et al. 2019)

Sources: Braga, A. A., Barao, L. M., Zimmerman, G. M., Douglas, S., and Sheppard, K. 2019. "Measuring the Direct and Spillover Effects of Body Worn Cameras on the Civility of Police-Citizen Encounters and Police Work Activities." *Journal of Quantitative Criminology* (https://doi.org/10.1007/s10940-019-09434-9).

Lum, Cynthia, Christopher S. Koper, David B. Wilson, Megan Stoltz, Michael Goodier, Elizabeth Eggins, Angela Higginson, and Lorraine Mazerolle. 2020. "Body-Worn Cameras' Effects on Police Officers and Citizen Behavior: A Systematic Review." *Campbell Systematic Reviews* 16:e1112 (https://doi.org/10.1002/cl2.1112).

Exploratory Research

Qualitative research is often exploratory and, hence, inductive: The researchers begin by observing social interaction or interviewing social actors in depth and then developing an explanation for what has been found. The researchers often ask questions such as "What is going on here?" "How do people interpret these experiences?" or "Why do people do what they do?" Rather than testing a hypothesis, the researchers are trying to make sense of some social phenomenon. They may even put off formulating a research question until after they begin to collect data—the idea is to let the question emerge from the situation itself (Brewer and Hunter 1989:54–58).

Black Youths' Strategies for Navigating Police Contact

How do young Black people attempt to shape encounters with police officers? Black youth are more likely than others to be the targets of aggressive policing, and Brittany N. Fox-Williams (2019) used exploratory research methods to begin to understand how they respond to encounters with the police rather than how police respond to them. For this purpose, Fox-Williams recruited as her interviewees 19 young Black New York City residents who were identified as "on track," meaning they were in school or working and committed to mainstream goals like a college degree and a middle-class occupation. Interviews occurred either in small groups that lasted one to one and one-half hours or in individual interviews for about an hour; some participated in both group and individual interviews. Her open-ended questions

allowed participants to discuss their strategies for interacting with police in their own words. Fox-Williams then reviewed the interview transcripts carefully and identified major themes that emerged in the comments.

Three of the most common strategies youth used in their encounters with police that Fox-Williams identified were avoidance, management, and symbolic resistance. The following quote describes a young man's use of an avoidance strategy (Fox-Williams 2019:124).

> "I'm not thinking anything, just keep walking. . . . I just feel like a robot, no emotion. . . . My mind is really clear in that moment because I don't want to look suspicious."

The Fox-Williams (2019) example illustrates how qualitative data can enrich understanding of the social world by revealing how participants make sense of and manage their social experiences. It also provides another perspective on the negative consequences of aggressive policing.

Explanations developed inductively from qualitative research can feel authentic because we have heard what people have to say in their own words and we have tried to see the social world as they see it. Explanations derived from qualitative research will be richer and more finely textured than they often are in quantitative research, but they are likely to be based on fewer cases from a limited area. We cannot assume the people studied in this setting are like others or that other researchers will develop explanations similar to ours to make sense of what was observed or heard. Because we do not initially set up a test of a hypothesis according to some specific rules, another researcher cannot come along and conduct the same test.

Descriptive Research

You learned in Chapter 1 that some social research is purely descriptive. Such research may not itself connect theory and data, but it is still a part of the research circle—it begins with data and proceeds only to the stage of making empirical generalizations based on those data (refer to Figure 2.10).

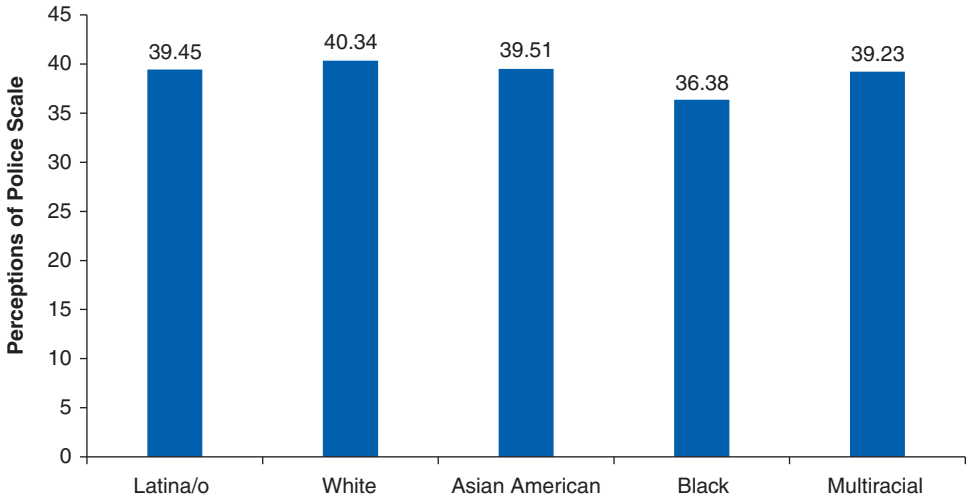
Valid description is important in its own right—and it is a necessary component of all investigations. Government agencies and nonprofit organizations frequently sponsor research that is primarily descriptive: How many poor people live in this community? Is the health of the elderly improving? How frequently do convicted criminals return to crime? Simply put, a good description of data is the cornerstone of the scientific research process and an essential component for understanding the social world.

Nadal et al. (2017) used both quantitative and qualitative methods to develop a description of perceptions of the police and to examine whether these perceptions differed between racial/ethnic groups and by age and gender. They recruited for a brief survey both students in a large urban minority-serving college and general community members. They then asked a subset of these respondents some questions about their own experiences with police.

The survey identified Black respondents as having a markedly less positive perception of the police than respondents in the other groups (Figure 2.15). The qualitative answers helped to explain why. For example, an 18-year-old Black man said, "The police told me they stopped me because I looked suspicious."

Good descriptive research can also stimulate more ambitious deductive and inductive research. Nadal et al. (2017:827) suggest the need "for more research to focus on how community policing affects different racial groups."

FIGURE 2.15 ■ Perceptions of Police by Race and Gender



Source: Based on Nadal, Kevin L., Kristin C. Davidoff, Neil Allicock, Chrine R. Serpe, and Tanya Erazo. 2017. "Perceptions of Police, Racial Profiling, and Psychological Outcomes: A Mixed Methodological Study." *Journal of Social Issues* 73(4):808-830.

SOCIAL RESEARCH STANDARDS

Social science research can improve our understanding of empirical reality—the reality we encounter firsthand. We have achieved the goal of **validity** when our conclusions about this empirical reality are correct. I look out my window and observe that it is raining—a valid observation, if my eyes and ears are to be trusted. I pick up the newspaper and read that the rate of violence may be climbing after several years of decline. I am less certain of the validity of this statement, based as it is on an interpretation of some trends in crime indicators obtained through some process that isn't explained. As you learned in this chapter, many social scientists who have studied police interaction with citizens came to the conclusion that there are racial biases in police shootings and also that body-worn cameras can reduce the frequency of police–citizen interactions that produce citizen complaints. However, there are many factors that affect the impact of BWCs, and more research is needed to guide public policies that seek to use BWCs to reduce problems due to aggressive policing.

If validity sounds desirable to you, you're a good candidate for becoming a social scientist. If you recognize that validity is often a difficult goal to achieve, you may be tough enough for social research. In any case, the goal of social science is not to come up with conclusions that people will like or conclusions that suit our own personal preferences. The goal is to figure out how and why the social world—some aspect of it—operates as it does. In *Investigating the Social World*, we are concerned with three standards for validity: (1) **measurement validity**, (2) **generalizability**, and (3) **causal validity** (also known as **internal validity**) (Hammersley 2008:43). We will learn that invalid measures, invalid generalizations, or invalid causal inferences will result in invalid conclusions. We will also focus on the standard of **authenticity**, a concern with reflecting fairly the perspectives of participants in a setting that we study.

Measurement Validity

Measurement validity is our first concern in establishing the validity of research results because without having measured what we think we measured, we really don't know what we're talking about. Measurement validity is the focus of Chapter 4. A measure is valid when it measures what we think it measures. In other words, if we seek to describe the frequency of domestic violence in families, we need to develop a valid procedure for measuring domestic violence.

The first step in achieving measurement validity is to specify clearly what it is we intend to measure. Patricia Tjaden and Nancy Thoennes (2000) identified this as one of the problems in the area of research on domestic violence: "definitions of the term vary widely from study to study, making comparisons difficult" (p. 5). To avoid this problem, Tjaden and Thoennes (2000) presented a clear definition of what they meant by *intimate partner violence:*

> Rape, physical assault, and stalking perpetrated by current and former dates, spouses, and cohabiting partners, with cohabiting meaning living together at least some of the time as a couple. (p. 5)

Tjaden and Thoennes also provided a measure of each type of violence. For example, "'physical assault' is defined as behaviors that threaten, attempt, or actually inflict physical harm" (Tjaden and Thoennes 2000:5). With this definition in mind, Tjaden and Thoennes (2000:6) then specified the set of questions they would use to measure intimate partner violence (the questions pertaining to physical assault):

> Not counting any incidents you have already mentioned, after you became an adult, did any other adult, male or female, ever:
>
> - Throw something at you that could hurt?
> - Push, grab, or shove you?
> - Pull your hair?
> - Slap or hit you?
> - Kick or bite you?
> - Choke or attempt to drown you?
> - Hit you with some object?
> - Beat you up?
> - Threaten you with a gun?
> - Threaten you with a knife or other weapon?
> - Use a gun on you?
> - Use a knife or other weapon on you?

Do you believe that answers to these questions provide a valid measure of having been physically assaulted? Do you worry that some survey respondents might not report all the assaults they have experienced? Might some respondents make up some incidents? Issues like

these must be considered when we evaluate measurement validity. Suffice it to say that we must be very careful in designing our measures and in subsequently evaluating how well they have performed. Chapter 4 introduces several different ways to test measurement validity. We cannot just *assume* that measures are valid.

Generalizability

The generalizability of a study is the extent to which it can be used to inform us about persons, places, or events that were not studied. Generalizability is the focus of Chapter 5.

You have already learned in this chapter that the Braga et al. (2018) findings about the effect of BWCs on police behavior did not hold up in some studies in other locations: As you know, this led to additional research to figure out what accounted for the different patterns in different locations.

If every person or community we study were like every other one, generalizations based on observations of a small number would be valid. But that's not the case. We are on solid ground if we question the generalizability of statements about research based on the results of a restricted sample of the population or in just one community or other social context.

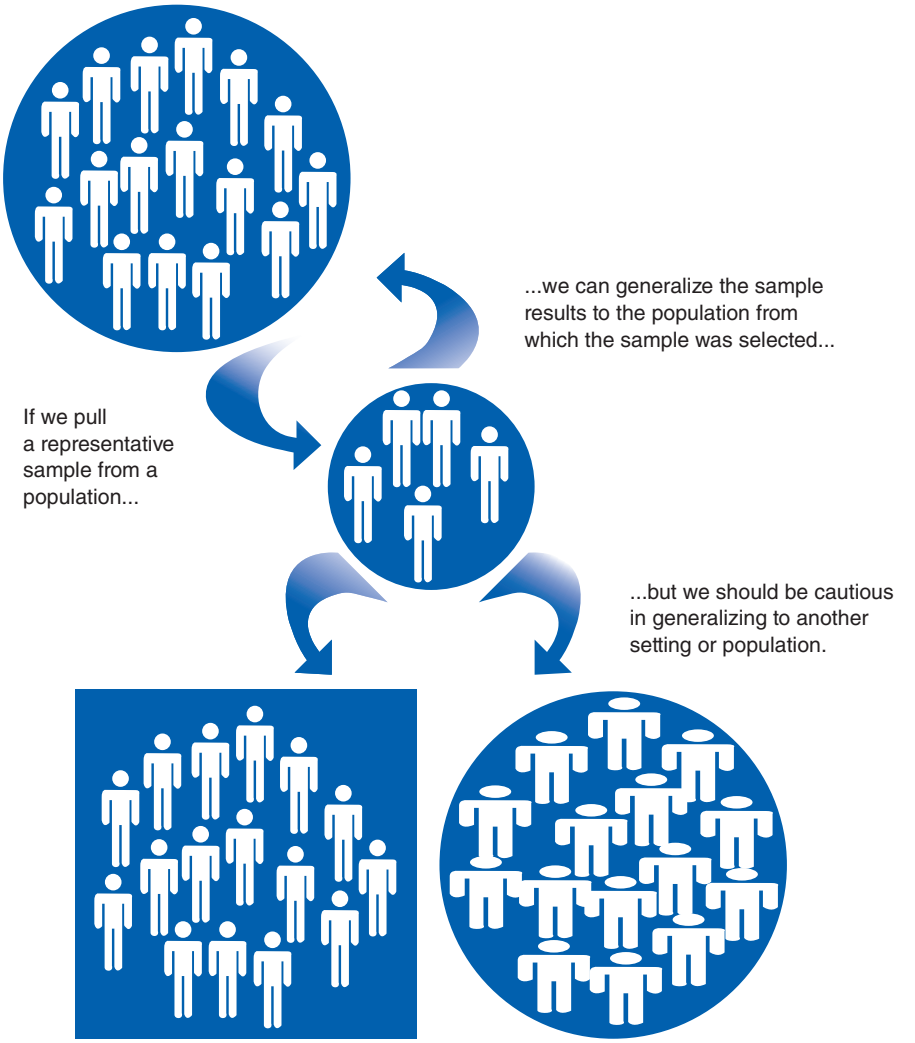
Generalizability has two aspects. **Sample generalizability** refers to the ability to generalize from a sample, or subset, of a larger population to that population itself. This is the most common meaning of generalizability. **Cross-population generalizability** refers to the ability to generalize from findings about one group, population, or setting to other groups, populations, or settings (see Figure 2.16). Cross-population generalizability can also be referred to as **external validity**. (Some social scientists equate the term *external validity* to *generalizability*, but in this book, I restrict its use to the more limited notion of cross-population generalizability.)

Sample generalizability is a key concern in survey research. Political pollsters may study a sample of likely voters, for example, and then generalize their findings to the entire population of likely voters. No one would be interested in the results of political polls if they represented only the relatively tiny sample that actually was surveyed rather than the entire population. The procedures for the National Violence Against Women Survey that Tjaden and Thoennes (2000) relied on were designed to maximize sample generalizability.

Cross-population generalizability occurs to the extent that the results of a study hold true for multiple populations; these populations may not all have been sampled or they may be represented as subgroups within the sample studied. This was the problem that arose in relation to a study that Sherman and Berk (1984) conducted of the impact of a mandatory arrest policy for police responding to calls about domestic violence in Minneapolis. Persons arrested for the same crime in several other locations in subsequent studies did not respond in the same way. The conclusions from Sherman and Berk's (1984) initial research in Minneapolis were not "externally valid."

Generalizability is a key concern in research design. We rarely have the resources to study the entire population that is of interest to us, so we have to select cases to study that will allow our findings to be generalized to the population of interest. Chapter 5 reviews alternative approaches to selecting cases so that findings can be generalized to the population from which the cases were selected. Nonetheless, because we can never be sure that our findings will hold under all conditions, we should be cautious in generalizing to populations or periods that we did not sample.

FIGURE 2.16 ■ Sample and Cross-Population Generalizability


...we can generalize the sample results to the population from which the sample was selected...
If we pull a representative sample from a population...
...but we should be cautious in generalizing to another setting or population.

Causal Validity

Causal validity, also known as *internal validity*, refers to the truthfulness of an assertion that A causes B. It is the focus of Chapter 6.

Most research seeks to determine what causes what, so social scientists frequently must be concerned with causal validity. Braga et al. (2018) were concerned with the effect of BWCs on the likelihood of citizen complaints about police actions. To test their causal hypothesis, Braga et al. designed their experiment so that some police responding to incident reports wore body cameras while some responding to other incident reports did not. Of course, it may seem heavy-handed for social scientists to influence police actions for the purpose of a research project, but this step reflects just how difficult it can be to establish causally valid understandings about the social world. Only because police did not get to choose whether they would wear a body camera when responding to an incident can we be reasonably sure that the officers wearing body cameras and the incidents they were responding to were very similar to the officers who were *not* wearing body cameras and the incidents they were responding to.

Chapter 6 will give you much more understanding of how some features of a research design can help us evaluate causal propositions. However, you will also learn that the solutions are neither easy nor perfect: We always have to consider critically the validity of causal statements that we hear or read.

Authenticity

The goal of authenticity is stressed by researchers who focus attention on the subjective dimension of the social world. An authentic understanding of a social process or social setting is one that reflects fairly the various perspectives of participants in that setting (Gubrium and Holstein 1997). Authenticity is one of several different standards proposed by some observers as uniquely suited to qualitative research; it reflects a belief that those who study the social world should focus first and foremost on how participants view that social world rather than on developing a unique social scientists' interpretation of that world. Rather than expecting social scientists to be able to provide a valid mirror of reality, this perspective emphasizes the need for recognizing that what is understood by participants as reality is a linguistic and social construction of reality (Kvale 2002:306).

Angela M. Moe (2007) explained her basis for considering the responses of women she interviewed in a domestic violence shelter to be authentic:

> Members of marginalized groups are better positioned than members of socially dominant groups to describe the ways in which the world is organized according to the oppressions they experience. (p. 682)

Moe's (2007) assumption was that "battered women serve as experts of their own lives" (p. 682). Adding to her assessment of authenticity, Moe (2007) found that the women "exhibited a great deal of comfort through their honesty and candor" as they produced "a richly detailed and descriptive set of narratives" (p. 683). You will learn more about how authenticity can be achieved in qualitative methods inChapters 10 and 11.

CONCLUSIONS

Selecting a worthy research question does not guarantee a worthwhile research project. The simplicity of the research circle presented in this chapter belies the complexity of the social research process. In the following chapters, I focus on particular aspects of the research process. Chapter 4 examines the interrelated processes of conceptualization and measurement, arguably the most important part of research. Measurement validity is the foundation for the other two aspects of validity.Chapter 5 reviews the meaning of generalizability and the sampling strategies that help us achieve this goal. Chapter 6 introduces causal validity and illustrates different methods for achieving it. Most of the remaining chapters then introduce different approaches to data collection—experiments, surveys, participant observation and intensive interviewing, evaluation research, comparative historical research, Big Data analysis, secondary data analysis, and content analysis—that help us, in different ways, achieve results that are valid.

Of course, our answers to research questions will never be complete or entirely certain. We always need to ground our research plans and results in the literature about related research. Our approach should be guided by explicit consideration of a larger theoretical framework. When we complete a research project, we should evaluate the confidence that can be placed in our conclusions, point out how the research could be extended, and consider the implications for social theory. Recall how the elaboration of knowledge about deterrence of domestic

violence required sensitivity to research difficulties, careful weighing of the evidence, identification of unanswered questions, and consideration of alternative theories.

If you will conduct your own research, consider keeping a research journal in which you keep track of your decisions, question yourself at each step, and reflect on what you are reading and experiencing (Ravitch and Riggan 2016:216). It will help a lot when it comes time to write up the methods you have used.

Owning a large social science toolkit and even working with a good research team is no guarantee for making the right decisions about which tools to use and how to use them in the investigation of particular research problems. However, you are now forewarned about and thus I hope forearmed against some of the problems that social scientists face in their work. I hope that you will return often to this chapter as you read the subsequent chapters, when you criticize the research literature, and when you design your own research projects. To be conscientious, thoughtful, and responsible—this is the mandate of every social scientist. If you formulate a feasible research problem, ask the right questions in advance, try to adhere to the research guidelines, and steer clear of the most common difficulties, you will be well along the road to fulfilling this mandate.

KEY TERMS

Anomalous findings (p. 62)
Authenticity (p. 66)
Causal validity (internal validity)
Conflict theory (p. 39)
Cross-population generalizability (external validity)
Deductive research (p. 58)
Dependent variable (p. 58)
Direction of association (p. 59)
Empirical generalization
External validity (cross-population generalizability)
Functionalism (p. 39)
Generalizability (p. 66)
Hypothesis (p. 58)
Independent variable (p. 58)
Inductive research (p. 62)
Internal validity (causal validity)
Measurement validity (p. 66)
Normal science (p. 40)
Rational choice theory (p. 38)
Research circle (p. 58)
Sample generalizability (p. 68)
Scientific paradigm
Serendipitous findings (p. 62)
Social research question (p. 33)
Symbolic interaction theory (p. 38)
Systematic review (p. 51)
Theory (p. 37)
Validity (p. 66)
Variable

HIGHLIGHTS

LO 2.1 Appraise social research questions based on their feasibility, social importance, and scientific relevance.

- Social research questions should be feasible in that one is able to conduct a study to answer them with the time and resources available and gain access to the group one desires to study.
- Social research questions should have a focus on a substantive area of importance.
- Social research questions should be scientifically relevant and grounded in the social science literature.

LO 2.2 Compare the concepts of "theory" and "paradigm."

- A theory is a logically interrelated set of propositions that help us make sense of many interrelated phenomena and predict behavior or attitudes that are likely to occur under certain conditions.
- A scientific paradigm is a set of beliefs that guide scientific work in an area, including unquestioned presuppositions, accepted theories, and exemplary research findings.

LO 2.3 Devise a strategy for searching the literature and the web about a research question.

- Choose bibliographic databases for your search that compile references from the peer-reviewed journals in your area(s) of interest.
- Identify your research question, appropriate bibliographic databases to search, and a tentative list of search terms.
- Narrow your search using Boolean search logic and appropriate subject descriptors.
- Locate the articles.
- Search the web for related research reports and other unpublished sources, research project websites, and other material relevant to your research question, but be careful to focus on credible sources and to limit your search to the most relevant topics.

LO 2.4 Design a literature review for a research question.

Reviewing peer-reviewed journal articles that report prior research is an essential step in designing new research.

- Summarize each selected article in a single-article review.
- Integrate the article reviews to highlight the implications of the articles for different aspects of your research question and procedures.
- Seek systematic reviews that have integrated the findings on your research question.

LO 2.5 Assess when a research project is designed to answer primarily a deductive or an inductive research question.

- Research based on deductive reasoning proceeds from general ideas, deduces specific expectations from these ideas, and then tests the ideas with empirical data.
- Research based on inductive reasoning begins with specific data and then develops general ideas or theories to explain patterns in the data.

LO 2.6 Compare the four standards for social research: measurement validity, generalizability, causal validity, and authenticity.

- Measurement validity exists when a measure measures the thing we think it measures.
- There are two aspects of generalizability. Sample generalizability refers to the ability to generalize from a subset of a population (a sample) to the entire population. Cross-population generalizability (external validity) refers to the ability to apply findings about one population to other populations.
- Causal (internal) validity refers to the truthfulness of an assertion that A causes B.
- Authenticity means reflecting fairly the various perspectives of participants in that setting.

DISCUSSION QUESTIONS

1. Pick a social issue about which you think research is needed. Draft three research questions about this issue. Refine one of the questions and evaluate it in terms of the three criteria for good research questions.
2. Identify variables that are relevant to your three research questions. Now formulate three related hypotheses. Which are the independent and which are the dependent variables in these hypotheses?
3. If you were to design research about police–citizen interaction, would you prefer an inductive approach or a deductive approach? Explain your preference. What would be the advantages and disadvantages of each approach? Consider in your answer the role of social theory, the value of searching the literature, and the goals of your research.
4. Braga et al.'s (2020) study of the effect of body-worn cameras tested a prediction derived from rational choice theory. Propose hypotheses about the effect of body-worn cameras that are consistent with conflict and symbolic interactionist theories. Which theory seems to you to provide the best framework for understanding the effect of body-worn cameras?
5. Review my description of the research projects in the section "Types of Social Research" inChapter 1. Can you identify the stages of each project corresponding to the points on the research circle? Did each project include each of the four stages? Which theory (or theories) seems applicable to each of these projects?
6. Fox-Williams (2019) studied responses of Black youth to encounters with the police using an exploratory research approach. Why do you think the researcher adopted this approach? Do you agree with their decision? Propose a research project that would address issues in one of these studies with a deductive approach.
7. Critique the Braga et al. (2020) research on the effect of body-worn cameras from the standpoint of measurement validity, generalizability, and causal validity. What else would you like to know about this research so that you can strengthen your critique? What does consideration of the goal of authenticity add to your critique?

SPSS EXERCISES

1. Describe beliefs about police.

   After opening the GSS2020 data file, click Analyze/Descriptive Statistics/Frequencies and find the variable poltrtblk_2. (Right click on the list of variables and tell it to display variable names instead of variable labels and sort alphabetically.) Then, move it into the variable list and click OK. Examine the output file. To what extent would you say people feel police treat white people better than Black people?

   After opening the GSS2020 data file, click Analyze/Descriptive Statistics/Frequencies and find the variable defund_2. (Right click on the list of variables and tell it to display variable names instead of variable labels and sort alphabetically.) Then, move it into the variable list and click OK. Examine the output file. What percentage of the sample favored defunding the police?

2. Test 2 hypotheses that predict support for defunding the police. Let's try education level and region of the country. Would you hypothesize that support for defunding the police would increase or decrease with educational level? Would you hypothesize that support for defunding the police would be higher in some regions of the United States than in others? Write out your hypotheses and then go on to steps a. and b. to test them.
   - **a.** By degree—How did support for defunding vary by education level?
     - **i.** Click on Analyze/Descriptive Statistics/Crosstabs.
     - **ii.** Highlight defund_2 and click on the arrow so that it moves into the Rows box; highlight degree_2 and click on the arrow to move it into the Columns box.
     - **iii.** Click on Cells, click off Counts-Observed, and click on Percentages-Column.
     - **iv.** Click Continue and then OK. Inspect the table.
   - **b.** By region—How did support for defunding vary by region?
     - **i.** Click on Analyze/Descriptive Statistics/Crosstabs.
     - **ii.** Highlight defund_2 and click on the arrow so that it moves into the Rows box; highlight region_2 and click on the arrow to move it into the Columns box.
     - **iii.** Click on Cells, click off Counts-Observed, and click on Percentages-Column.
     - **iv.** Click Continue and then OK. Inspect the table.

Tuskegee Syphilis Study Administrative Records. Records of the Centers for Disease Control and Prevention. National Archives—Southeast Region (Atlanta).

3 RESEARCH ETHICS AND RESEARCH PROPOSALS

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

3.1 Discuss the human rights violations that led to the Belmont Report and subsequent regulations

3.2 Judge the extent to which a research design satisfies each of the ethical principles for the protection of human subjects

3.3 Compose an informed consent form

3.4 Defend the institutional review board requirement for informed consent in a research study

3.5 Create a detailed outline for each section of a research proposal

Can you imagine being a tenant farmer during the Great Depression? Tenant farming and the related system of sharecropping replaced slavery in much of the South after the Civil War, and it never made for an easy way to live. Tenant farmers worked a plot of land—often part of a former plantation—and then paid the landowner for using the land at the end of the season with proceeds from the cotton they sold. Contracts often kept tenant farmers and sharecroppers "in an almost inescapable cycle of poverty," perpetually in debt to the landowner and to local businesses from which they purchased food and supplies (Ferguson 1998:38). Beatings and lynchings were common. Meager earnings—a median of just $1 per day—were often insufficient for necessities, to say nothing of "luxuries" like professional medical care. By the end of 1932, the third year of the Depression, almost one quarter of the U.S. labor force was unemployed, and any hope you may have had for a better life would most likely have vanished.

So how would you have reacted if you learned that medical professionals would provide free medical care to hundreds of people in your area who agreed to some health assessments and screening tests? If you were like 4,400 poor Black tenant farmers who received this offer in Macon County, Georgia, in 1932, you would have been grateful for the opportunity and turned out for the tests (Benedek 1978). Even though completion of the enrollment process required a follow-up visit for a painful spinal tap, many stuck it out, and the program ultimately accepted 399 of the men into the "free medical care" program. Most had never even seen a doctor before nor been able to afford medicines, so becoming eligible for medical exams and medications like aspirin and vitamins at no cost seemed to be quite a stroke of luck.

And in fact, that was how Macon County's new health program had initially been conceived in 1898 by Booker T. Washington, founder of the Tuskegee Institute in Macon County, and Julius Rosenwald, founder of Sears, Roebuck & Co. and a leading philanthropist (Washington 2006:158–160). Unfortunately, Rosenwald lost his wealth in the 1929 stock market crash that began the Great Depression, and so funds for launching the Macon County medical program vanished. The U.S. Public Health Service (PHS) then agreed to support the program, but the chief of the PHS Venereal Disease Division, Dr. Taliaferro Clark, decided it would now be used to study the course of untreated syphilis rather than to improve the health of program participants. Reflecting on high rates of syphilis in the local population, one PHS doctor even imagined a potential benefit for

white racists rather than for the study's participants, as "the scourge sweeps among" Black Alabamans and their untreated syphilis results in prolonged illness and death (Washington 2006:160):

Perhaps here, in conjunction with tuberculosis, will be the end of the negro problem. Disease will accomplish what man cannot do.

So what had been conceived as an effort to improve the health of poor Black people in the South became a program taking advantage of their impoverished circumstances and lack of health care to better understand the course of a dreaded venereal disease and ultimately improve treatment—for others who could afford it. Despite this change in purpose, the program continued to attract participants with such announcements as, "YOUR LAST CHANCE FOR SPECIAL FREE TREATMENT" (Washington 2006:161). Unbeknown to those who enrolled, the subsequent medical tests and spinal taps were used to identify those with syphilis, not to determine a course of treatment (see Figure 3.1). Thus began the Tuskegee Study of Untreated Syphilis in the Negro Male. The U.S. Public Health Service study of the "natural" course of syphilis followed 399 low-income African American men from the 1930s to 1972. (Washington 2006:157).

In 1933. O. C. Wenger, head of PHS syphilis programs, remarked that, "As I see it, we have no further interest in these patients until they die" (when their bodies could be autopsied). It would be hard to make up a statement that more boldly demonstrates how racism diminishes the humanity of those it targets.

For forty years, men in the study received no treatment, beyond occasional visits from a local nurse who supplied aspirin, vitamins, and rides into town for medical tests. The men were even exempted from the draft during World War II for fear they would receive treatment as servicemen. What they did receive was an offer for financial assistance for burials but without being informed that the offer was made only to ensure their dead bodies could first be autopsied. In 1943, penicillin was identified as a cure for syphilis, much more effective than previous medications, and the PHS helped to lead a national campaign to encourage its use. The Tuskegee participants, however, were intentionally not informed. In 1958, the PHS awarded those men still alive a certificate of appreciation and $25—one dollar for each year since the study began—even as the observational study continued without providing treatment (Washington 2006:163–169).

It was not until 1965 that Peter Buxton, a Polish immigrant working for PHS learned about the Tuskegee Study, considered it unethical, and urged PHS leaders to terminate it. It was not until 1972, after no changes were made in response to his complaint, that Buxton informed a journalist about the study and an exposé was published, followed by widespread condemnation of the study, a congressional hearing, and then a review by an ad hoc panel of the Department of Health, Education, and Welfare. The panel declared the study unethical, recommended its termination, and called for a new national committee to regulate human subjects research (Caplan 1992). Participants and their families began to receive treatment. In 1974, participants received a $10 million settlement after a class-action lawsuit ($37,500 per living participant and $15,000 for their heirs). In 1997, President Bill Clinton (1997) issued an official apology in a White House ceremony that included the remaining eight survivors and their families and granted funds to establish the Tuskegee University National Center for Bioethics in Research and Health Care.

In the early years of the Tuskegee Study and even in the decades after it was ended, some have argued that the lack of effective treatments for syphilis at the time the study began, lack of knowledge about the long-term course of syphilis, the unavailability

of treatment for most people and low levels of adherence to what was a very lengthy treatment protocol before the advent of penicillin, and the age of most participants justified the research design (Bedenek and Erlen 1999; White 2000). But the failure to allow the participants to weigh these issues themselves and the lack of attention to the impact of the study on the participants' lives highlights the need to review research procedures for adherence to ethical standards and to avoid simply assuming that researchers' motives are beneficent. Publicity about the study drew attention to the importance of obtaining informed consent from research participants and encouraged the creation of the federal regulations that now guide all research involving human subjects. Understandably, this publicity also contributed to widespread distrust of the medical establishment and healthcare research in the Black community (Frazier 2020). We will learn more about current consequences of "Tuskegee" and other unethical research conducted on Black Americans in later sections of this chapter.

HISTORICAL BACKGROUND

Concern with ethical practice in relation to people who are in some respect dependent, whether as patients or research subjects, did not begin with the Tuskegee study. Ethical guidelines for medicine trace back to Hippocrates in 5 BC Greece (Hippocratic Oath n.d.), and the American Medical Association (AMA) adopted the world's first formal professional ethics code in medicine in 1847 (AMA 2011). Current AMA ethical principles include respecting patient rights, maintaining confidentiality, and regarding "responsibility to the patient as paramount" (AMA 2011). Yet the history of medical practice makes it clear that having an ethics code is not sufficient to ensure ethical practice, at least when there are clear incentives to do otherwise.

A defining event occurred in 1946, when the **Nuremberg War Crime Trials** exposed the horrific medical experiments conducted by Nazi doctors and others in the name of "science." When considering the need for ethical guidelines in research, it is sobering to realize that what became a famous example of ethically questionable social science research in the 1960s was motivated in part by a desire to understand why so many Germans seemed to support the Nazis.

Maybe you will find it easier to imagine yourself reading a newspaper ad in New Haven, Connecticut, in 1960 than living as a tenant farmer in Georgia in 1932. How do you think you would have responded to an advertisement offering an opportunity to "Earn money and learn about yourself" by participating in a psychology experiment at Yale University?

WE WILL PAY YOU $45 FOR ONE HOUR OF YOUR TIME

Persons Needed for a Study of Memory

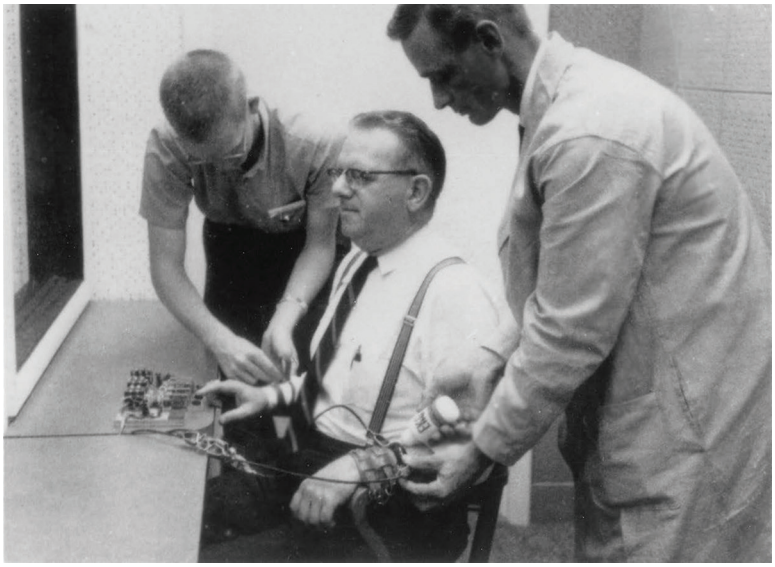
If you then checked into the designated waiting room at Yale, a man dressed in a lab coat would have introduced himself as a psychologist interested in whether people learn things better when they are punished for making a mistake.

"The experimenter" (as we'll refer to him from now on) says he will write either *teacher* or *learner* on small identical slips of paper and then asks you and another person in the room to draw out one. Yours says *teacher*.

The experimenter now states, in a matter-of-fact way, "All right. Now the first thing we'll have to do is to set the learner up so that he can get some type of punishment."

He leads you both behind a curtain, sits the learner down, attaches a wire to his left wrist, and straps both his arms to the chair so he cannot remove the wire (see Figure 3.1). The wire is connected across the room to a console with 30 switches and a large dial. The experimenter asks

FIGURE 3.1 ■ Learner Strapped in Chair With Electrodes



Source: From the film *OBEDIENCE*. Copyright © 1968 by Stanley Milgram, copyright renewed 1993 by Alexandra Milgram and distributed by Alexander Street Press.

you to hold the end of the wire, walks back to the control console, flips several switches, and you hear a clicking noise. The dial moves, and you feel an electric shock. The experimenter explains the machine is calibrated so it will not cause permanent injury, but he also acknowledges that when it is turned up all the way, it is very, very painful and can result in severe, although momentary, discomfort.

Now you walk back to the other side of the room (so the learner is behind the curtain) and sit before the console. The experimental procedure is simple: (1) You read aloud a series of word pairs, such as *blue box, nice day, wild duck*, and so on. (2) You then read one of the first words from those pairs and a set of four words, one of which contains the original paired word. (3) If the learner states the word he thinks was paired with the first word you read, you are to compliment him and move on to the next word. If he makes a mistake, you flip a switch on the console and the learner feels a shock. (4) After each mistake, you are to flip the next switch on the console, progressing from a mark labeled *slight shock* on up to marks labeled *intense shock, extreme intensity shock*, and *danger: severe shock.*

You begin. As you turn the dial, the learner's responses increase in intensity from a grunt at the tenth mark (strong shock) to painful groans at higher levels, anguished cries to "get me out of here" at the extreme intensity shock levels, to a deathly silence at the highest level. If you indicate discomfort at administering the stronger shocks, the experimenter tells you, "The experiment requires that you continue." Now, please note on the meter in Figure 3.2 the most severe shock you would agree to give to the learner.

You may recognize this thought experiment is a slightly simplified version of **Milgram's obedience experiments**. The 40 New Haven adults who volunteered for the experiment at Yale University in the early 1960s administered an average level of shock of 24.53, or a level close to the label "extreme intensity shock" and almost in the "red zone" indicating "Danger: severe shock." Almost two thirds complied with the experimenter's demands all the way to the top of the scale (originally labeled simply as *XXX,* then indicated numerically as *30*).

FIGURE 3.2 ■ Shock Meter

There is abundant evidence from the subjects' own observed high stress and their subsequent reports that many subjects really believed the learner was receiving actual, hurtful shocks. Are you surprised by the subjects' responses? Do you think the results of this experiment tell us about how people behave in the real world? Can you imagine why Milgram's (1963) research on obedience ultimately had a profound influence on the way social scientists think about research ethics as well as on the way they understand obedience to authority? Like exposure of the Tuskegee research, the controversy over the ethics of the Milgram research fueled efforts to ensure the protection of human subjects in research.

Stanley Milgram's research on obedience generated controversy about participant protections (Perry 2013:37). Harvard medical researcher Henry K. Beecher published in 1966 an article in the prestigious *New England Journal of Medicine* that described 22 unethical experimental studies of human subjects, usually without their knowledge (Israel 2014:32–33). In 1970, sociologist Laud Humphreys's book, *Tearoom Trade*, revealed his secretive observations of men engaged in sex with other men in public bathrooms, followed by his interviewing them in their homes under false pretenses (described later in this chapter); and in 1971, sociologist Philip Zimbardo conducted his famous prison simulation experiment at Stanford University to learn how ordinary people would treat each other if they took roles of being prisoners and guards (the answer: not well) (Israel 2014:24–28). And then came the 1972 exposure of the Tuskegee study.

These and other widely publicized cases convinced many observers that formal review procedures were needed to protect research participants. As concerns about research ethics increased, the U.S. government created a National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research and charged it with developing guidelines (Kitchener and Kitchener 2009:7). The commission's 1979 ***Belmont Report*** (Department of Health, Education, and Welfare 1979) established three basic ethical principles for the protection of human subjects:

- **Respect for persons:** treating persons as autonomous agents and protecting those with diminished autonomy
- **Beneficence:** minimizing possible harms and maximizing benefits
- **Justice:** distributing benefits and risks of research fairly

The U.S. Department of Health and Human Services (DHHS) and the U.S. Food and Drug Administration then translated these principles into specific regulations that were adopted in 1991 as the **Federal Policy for the Protection of Human Subjects**, also known as the **Common Rule** (Title 45 of Code of Federal Regulations [CFR], Part 46). This policy has shaped the course of social science research ever since, by requiring organizations that sponsor federally funded research—including universities— to establish committees that review all research proposed at the institution and ensure compliance with the federal human subjects requirements when the research is conducted. Professional associations, such as the American Sociological Association (ASA), institutional review boards (IRBs) at universities and other research sponsors, and ethics committees in some organizations, also set standards for the treatment of human subjects by their members, employees, and students, although these standards are all designed to comply with the federal policy.

After 25 years, the Common Rule was revised and the changes required in the **Revised Common Rule** were implemented in 2019 (Menikoff, Kaneshiro, and Pritchard 2017; Revised Common Rule Q&As | HHS.gov). The revisions relaxed some requirements for social science research and made several other important changes that are consequential for both medical and social science researchers. These regulations, as revised, inform the discussion that follows (Baumann 2017; Chadwick 2017).

ETHICAL PRINCIPLES

The ASA, like other professional social science organizations, has adopted, for practicing sociologists, ethical guidelines that are more specific than the federal regulations. Professional organizations may also review complaints of unethical practices when asked.

The **Code of Ethics** of the ASA (1999) is summarized on the ASA website (http://www.asanet.org/membership/code-ethics); the complete text of the code is also available at this site. The general principles articulated in the code are intended to guide professional practice in diverse settings while some of the specific standards focus specifically on the protection of human subjects in research.

According to the general principles, sociologists should be committed in their work to high levels of competence, to practicing with integrity, and to maintaining responsibility for their actions. They must also respect the rights, dignity, and diversity of others, including research participants, as well as be socially responsible to their communities and use research to contribute to the public good. The following sections discuss the most important implications of these principles for the conduct of research, including the protection of human subjects.

Achievement of Valid Results

Commitment to achieving valid results is the necessary starting point for ethical research practice. Simply put, we have no business asking people to answer questions, submit to observations, or participate in experimental procedures if we are simply seeking to verify our preexisting prejudices or convince others to take action on behalf of our personal interests. The pursuit of objective knowledge about human behavior—the goal of validity—motivates and justifies our investigations and gives us some claim to the right to influence others to participate in our research. Knowledge is the foundation of human progress as well as the basis for our expectation that we, as social scientists, can help people achieve a brighter future. If we approach our research projects objectively, setting aside our personal predilections in the service of learning a bit more

about human behavior, we can honestly represent our actions as potentially contributing to the advancement of knowledge.

Milgram made a strong case in his 1963 article and 1974 book on the obedience experiments that he was committed to achieving valid results—to learning how and why obedience influences behavior. He tied his motivations directly to the horror of the Holocaust, to which the world's attention had been drawn once again by the capture and trial of Adolf Hitler's mastermind of that genocide, Adolf Eichmann (Perry 2013:210). In Milgram's (1963) own words,

> It has been reliably established that from 1933–45 millions of innocent persons were systematically slaughtered on command. . . . Obedience is the psychological mechanism that links individual action to political purpose. It is the dispositional cement that binds men to systems of authority . . . for many persons obedience may be a deeply ingrained behavior tendency. . . . Obedience may [also] be ennobling and educative and refer to acts of charity and kindness, as well as to destruction. (p. 371)

Milgram (1963) then explains how he devised experiments to study the process of obedience in a way that would seem realistic to the subjects and still allow "important variables to be manipulated at several points in the experiment" (p. 372). According to Milgram, every step in the experiment was carefully designed to ensure that subjects received identical stimuli and that their responses were measured carefully. The experiment's design also reflected what had become in the preceding 30 years a tradition in social psychology of laboratory experiments that used deception to create different believable conditions for participants (Perry 2013:31–35).

Milgram (1963:377) made every effort to convince readers that "the particular conditions" of his experiment created the conditions for achieving valid results. These particular conditions included the setting for the experiment at Yale University, its purported "worthy purpose" to advance knowledge about learning and memory, and the voluntary participation of the subject as well as of the learner—as far as the subject knew. Milgram then tested the importance of some of these "particular conditions" (e.g., the location at Yale) in replications of the basic experiment (Milgram 1965).

However, not all social scientists agreed that Milgram's approach could achieve valid results. Milgram's first article on the research, "Behavioral Study of Obedience," was published in 1963 in the *Journal of Abnormal and Social Psychology.* In the next year, the *American Psychologist* published a critique of the experiment's methods and ethics by the psychologist Diana Baumrind (1964). Her critique begins with a rejection of the external validity—the generalizability—of the experiment, because

> The laboratory is unfamiliar as a setting and the rules of behavior ambiguous. . . . Therefore, the laboratory is not the place to study degree of obedience or suggestibility, as a function of a particular experimental condition. [And so,] the parallel between authority-subordinate relationships in Hitler's Germany and in Milgram's laboratory is unclear. (pp. 421–423)

Milgram (1964) quickly published a rejoinder in which he disagreed with (among other things) the notion that it is inappropriate to study obedience in a laboratory setting: "A subject's obedience is no less problematical because it occurs within a social institution called the psychological experiment" (p. 850).

Milgram (1974:169–178) also argued in his later book that his experiment had been replicated in other places and settings with the same results, that there was considerable evidence the subjects had believed they actually were administering shocks, and the "essence" of his experimental manipulation—the request that subjects comply with a legitimate authority—was also

found in the dilemma faced by people in Nazi Germany, soldiers at the My Lai massacre in Vietnam, and even the cultists who drank poison in Jonestown, Guyana, at the command of their leader, Jim Jones (Miller 1986:182–183).

Baumrind (1985) was still not convinced. In a follow-up article in *American Psychologist*, she argued that "far from illuminating real life, as he claimed, Milgram in fact appeared to have constructed a set of conditions so internally inconsistent that they could not occur in real life" (p. 171). Although Milgram died in 1984, the controversy did not. A recent review of the transcripts and interviews with many participants raises additional concerns about the experiment's validity (Perry 2013). Milgram understated the "experimenter's" efforts to get the subjects to comply, he overstated the subjects' level of obedience, he never publicized one condition in which most subjects refused to give strong shocks when the "learner" was a friend, and he didn't acknowledge that even those classified as "obedient" were looking for a way to get out of the experiment. His claim that the results were replicated in similar experiments around the world was only partially true, and it seems clear from the transcripts and interviews that the aura created by the location at Yale University and the emphasis on a contribution to "science" influenced many participants.

Do you agree with Milgram's assumption that obedience could fruitfully be studied in the laboratory? Do you find merit in Baumrind's criticism? Are you troubled by the new evidence that Milgram may have presented his evidence selectively, to make his conclusions as convincing as possible? Will your evaluation of the ethics of Milgram's experiments be influenced by your answers to these questions? Should our ethical judgments differ depending on whether we decide that a study provides valid information about important social psychological processes? Should it matter that a 2005 replication of Milgram's experiment (with less severe "shocks") for ABC TV supported Milgram's conclusions (Perry 2013:275–279)?

Critics also identified multiple problems with the scientific procedures used in the Tuskegee study. Early in the study, Tuskegee researchers claimed they had found unique cardiovascular features of the disease in Black people, but as early as 1933, a panel of the American Heart Association found the Tuskegee researchers had misinterpreted their own data (Washington 2006). The Tuskegee researchers also violated the integrity of their own research design by transferring members of the control group who developed syphilis into the group of men who already had syphilis. Were the researchers able to separate their preconceptions about Black men and syphilis from their conduct of the research?

I can't answer these questions about the Milgram and the Tuskegee studies for you, but bear in mind that both Milgram and his strongest critic at the time, Baumrind, buttressed their ethical arguments with assertions about the validity (or invalidity) of the experimental results. The Tuskegee research led to 13 publications, but the flaws identified in the research led the assistant chief of the Public Health Service's Venereal Disease branch to conclude by 1970 that the study had contributed nothing to preventing, finding, or treating infectious syphilis (Jones 1993). It is hard to justify *any* risk for human subjects or even *any* expenditure of time and resources if our findings tell us nothing about human behavior. When the stakes are as high as in the Tuskegee research—life or death—research that raises those stakes cannot be justified.

Honesty and Openness

The scientific concern with validity requires, in turn, that scientists be open in disclosing their methods and honest in presenting their findings. In contrast, research distorted by political or personal pressures to find particular outcomes or to achieve the most marketable results is unlikely to be carried out in an honest and open fashion. To assess the validity of a researcher's

conclusions and the ethics of their procedures, you need to know exactly how the research was conducted. This means that articles or other reports must include a detailed methodology section, perhaps supplemented by appendixes containing the research instruments or websites or an address where more information can be obtained.

The Tuskegee study's ethical lapses included a failure to publish or even refer to a research protocol focused on specific research questions. The researchers told the men that they were in a treatment program, not a research study. Although there were many problems with the treatments available for treatment of syphilis in the 1930s, by the end of the 1940s the value of penicillin was widely recognized and yet the men were not informed or treated (Benedek 1978). An article published by the Tuskegee researchers after twenty years of "untreated syphilis in the male Negro" provides detailed medical statistics but no details of the study's overall design (Olansky et al. 1956).

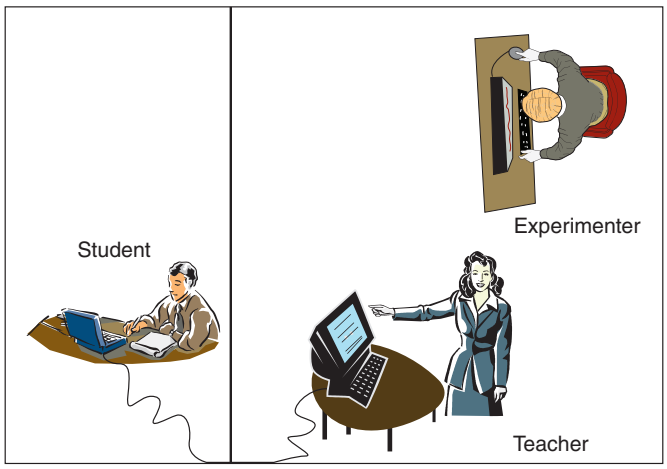
Milgram's initial 1963 article included a detailed description of study procedures that seemed to signal his adherence to the goal of honesty and openness. Many more details, including pictures, were provided in Milgram's (1974) subsequent book (see Figure 3.3).

The act of publication itself is a vital element in maintaining openness and honesty. Others can review and question study procedures and so generate an open dialogue with the researcher. Although Milgram disagreed sharply with Baumrind's criticisms of his experiments, their mutual commitment to public discourse in journals widely available to social scientists resulted in a more comprehensive presentation of study procedures and a more thoughtful discourse about research ethics. Almost 50 years later, this commentary continues to inform debates about research ethics (Cave and Holm 2003).

However, the latest significant publication in this open dialogue about Milgram's work calls into question his own commitment to the standard of openness and honesty. Gina Perry's (2013) *Behind the Shock Machine: The Untold Story of the Notorious Milgram Psychology Experiments* reveals many misleading statements by Milgram about participants' post-experiment debriefing, about adherence to the treatment protocol, about the extent of participants' apparent distress, and about the extent of support for his favored outcome.

Openness about research procedures and results thus goes hand in hand with honesty in research design and in research reporting. Despite this need for openness, some researchers may hesitate to disclose their procedures or results to prevent others from building on their ideas and

FIGURE 3.3 ■ Diagram of Milgram Experiment



Source: Northern Illinois University Department of Psychology.

taking some of the credit or, as may have occurred with Milgram, to make their procedures seem more acceptable or their findings more impressive. We just can't assume that everyone conducting research will avoid cutting corners in their work or failing to explain their procedures fully or even fabricating evidence if it helps to achieve their career goals. Just in 2011, Harvard psychology professor Marc Hauser resigned after an investigation concluded data he had reported in scientific publications was fabricated (Israel 2014:155). Cases like these demonstrate a lack of reflexivity, or researchers' failing to honestly consider how their own backgrounds, biases, and needs influence their research decisions (Poth 2021:79).

Conflicts of interest may occur when a researcher has a significant financial stake in the design or outcome of the research (American Sociological Association n.d.). Receiving speaking fees, consulting fees, patents or royalties, and other financial benefits from organizations that could be affected by research conclusions may lead researchers to distort their decisions and findings so as to protect these financial benefits. Many medical schools have regulations that prohibit their staff (including researchers) from accepting any tangible benefits from pharmaceutical companies for this reason. Both federal research funding agencies and journal editors require disclosure of possible conflicts of interest so others can scrutinize the extent to which these conflicts may have lessened researchers' honesty and openness (Fisher and Anushko 2008:96–97). Nonetheless, experimental research suggests that disclosure does not reduce trust in advice from experts even after they have disclosed a conflict of interest (Humphries 2011:K3).

Protection of Research Participants

Several standards concerning the treatment of human subjects are emphasized in federal regulations and the ethical guidelines adopted by many professional social science organizations:

- Research should cause no harm to subjects.
- Participation in research should be voluntary; therefore, subjects must give their informed consent to participate in the research, and researchers must disclose their identity.
- Researchers should avoid deception, except in limited circumstances.
- Anonymity or confidentiality must be maintained for individual research participants unless it is waived voluntarily and explicitly.
- Scientists must consider the uses of a research project so that its benefits outweigh any foreseeable risks.

Each of these standards became a focus of debate about Milgram's experiments, so we will return frequently to that debate as well to the Tuskegee legacy to keep our discussion realistic. We will also refer frequently to the ASA code to keep our treatment current. You will soon realize there is no simple answer to the question: What is (or isn't) ethical research practice? The issues are complicated and the relevant principles subject to different interpretations. However, by the time you finish this chapter, you will be aware of the major issues in research ethics and be able to make informed, defensible decisions about the ethical conduct of social science research.

Avoid Harming Research Participants

Although this standard may seem straightforward, it can be difficult to interpret in specific cases and harder yet to define in a way agreeable to all social scientists. Does it mean that subjects should not be harmed psychologically as well as physically at all? That they should feel no

anxiety or distress whatsoever during the study or only after their involvement ends? Should the possibility of any harm, no matter how remote, deter research?

The Tuskegee participants experienced multiple harms that the researchers could have avoided. Ten years after penicillin's adoption as the standard treatment for syphilis, the researchers noted in an article that "among the 23 living men who have developed some form of late syphilis, not one has received adequate therapy prior to the recent survey" without giving any reason for the failure of these patients to have received penicillin (Olansky et al. 1956:185). The successful effort by the researchers to ensure that none of the study participants received treatment for syphilis due to a Selective Service requirement in World War II prevented their referral for treatment at that point. Nor did the researchers seek to avoid the spread of disease to the men's wives and hence their children (Benedek 1978).

As we turn to Milgram's experiments, a verbatim transcript of one session will give you an idea of what participants experienced (Milgram 1965:67):

> **150 volts delivered.** You want me to keep going?
>
> **165 volts delivered.** That guy is hollering in there. There's a lot of them here. He's liable to have a heart condition. You want me to go on?
>
> **180 volts delivered.** He can't stand it! I'm not going to kill that man in there! You hear him hollering? He's hollering. He can't stand it. . . . I mean who is going to take responsibility if anything happens to that gentleman? [*The experimenter accepts responsibility.*] All right.
>
> **195 volts delivered.** You see he's hollering. Hear that. Gee, I don't know. [*The experimenter says: "The experiment requires that you go on."*] I know it does, sir, but I mean—hugh—he don't know what he's in for. He's up to 195 volts.
>
> **210 volts delivered.**
>
> **225 volts delivered.**
>
> **240 volts delivered.**

This experimental manipulation generated "extraordinary tension" (Milgram 1963:377):

> Subjects were observed to sweat, tremble, stutter, bite their lips, groan and dig their fingernails into their flesh. . . . Full-blown, uncontrollable seizures were observed for 3 subjects. One . . . seizure so violently convulsive that it was necessary to call a halt to the experiment [for that individual]. (p. 375)

An observer (behind a one-way mirror) reported, "I observed a mature and initially poised businessman enter the laboratory smiling and confident. Within 20 minutes he was reduced to a twitching, stuttering wreck, who was rapidly approaching a point of nervous collapse" (Milgram 1963:377).

From critic Baumrind's (1964:422) perspective, this emotional disturbance in subjects was "potentially harmful because it could easily effect an alteration in the subject's self-image or ability to trust adult authorities in the future." Milgram (1964) quickly countered,

> Momentary excitement is not the same as harm. As the experiment progressed there was no indication of injurious effects in the subjects; and as the subjects themselves strongly endorsed the experiment, the judgment I made was to continue the experiment. (p. 849)

When Milgram (1964:849) surveyed participants in a follow-up, 83.7% endorsed the statement they were "very glad" or "glad" "to have been in the experiment," 15.1% were "neither sorry

nor glad," and just 1.3% were "sorry" or "very sorry" to have participated (p. 849). Interviews by a psychiatrist a year later found no evidence "of any traumatic reactions" (p. 849)—although it was not disclosed that of 780 initial participants, only 140 were invited for an interview and only 32 of those accepted the invitation (Perry 2013:217). After these later revelations, Milgram's (1977:21) subsequent argument that "the central moral justification for allowing my experiment is that it was judged acceptable by those who took part in it" rings hollow.

Milgram (1963) also reported that he attempted to minimize harm to subjects with post-experimental procedures "to assure that the subject would leave the laboratory in a state of well being" (p. 374). He said that a friendly reconciliation was arranged between the subject and the victim and an effort was made to reduce any tensions that arose as a result of the experiment, but it turns out that his "dehoaxing" was typically very brief and did not disclose the deception to most participants. Most participants did not receive a letter informing them of the nature of the experiment until almost a year had passed (Milgram 1964:849; Perry 2013:72, 84).

Baumrind (1964:422) was unconvinced even without knowing of these later revelations: "It would be interesting to know what sort of procedures could dissipate the type of emotional disturbance just described [citing Milgram 1964]."

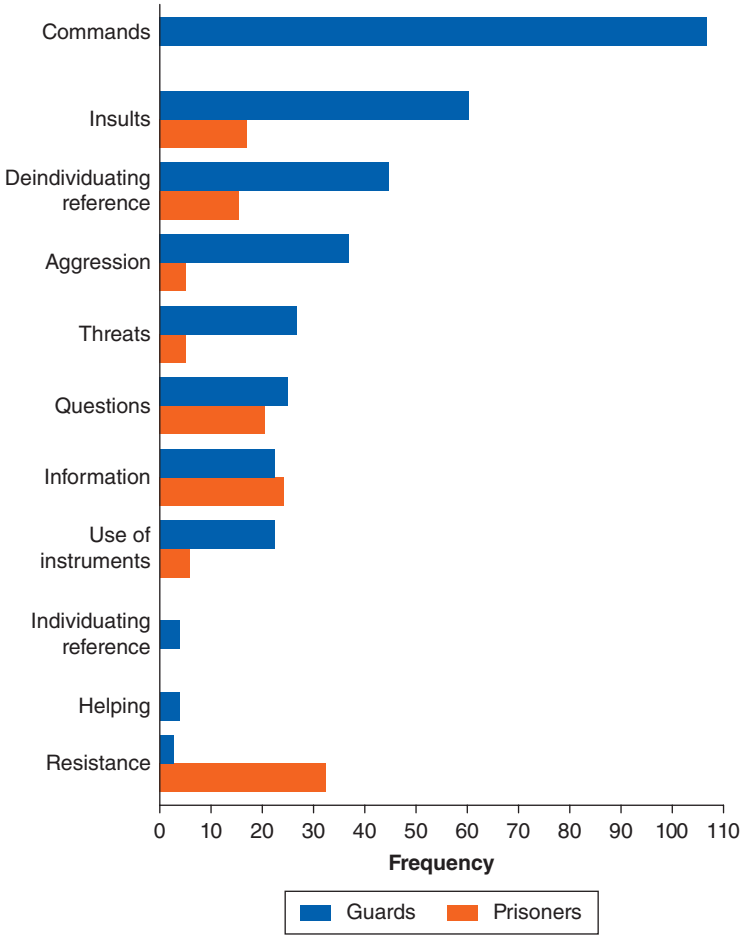
In a later article, Baumrind (1985:168) dismissed the value of the self-reported "lack of harm" to subjects who had been willing to participate in the experiment—although noting that still 16% did *not* endorse the statement that they were "glad" they had participated in the experiment. Baumrind (1985:169) also argued that research indicates most introductory psychology students (and some students in other social sciences) who have participated in a deception experiment report a decreased trust in authorities as a result—a tangible harm in itself.

Many social scientists, ethicists, and others concluded that Milgram's procedures had not harmed the subjects and so were justified for the knowledge they produced, but others sided with Baumrind's criticisms (Miller 1986:88–138; Perry 2013:269). Perry's (2013:77–78) recent investigation found even more evidence of psychological harm, including feelings of shame that had persisted since the experiment. The experimental records also reveal that debriefing never occurred for some participants and was very limited for almost all (Perry, 2013:76–84). Most were not told after the experiment that the shocks were fake; the usual "dehoaxing" consisted of the "learner" reassuring the "teacher" that the shocks he had received were not harmful.

What is your opinion of the possibility of harm at this point? Does Milgram's debriefing process relieve your concerns? Are you as persuaded by the subjects' own endorsement of the experiment as was Milgram?

What about possible harm to the subjects of the famous **Zimbardo prison simulation study** at Stanford University (Haney, Banks, and Zimbardo 1973)? The study was designed to investigate the impact of social position on behavior—specifically, the impact of being either a guard or a prisoner in a prison, a "total institution." The researchers selected apparently stable and mature young male volunteers and asked them to sign a contract to work for two weeks as a guard or a prisoner in a simulated prison. Within the first two days after the prisoners were incarcerated by the "guards" in a makeshift basement prison, the prisoners began to be passive and disorganized while the guards became "sadistic"—verbally and physically aggressive (see Figure 3.4). Five "prisoners" were soon released for depression, uncontrollable crying, fits of rage, and, in one case, a psychosomatic rash. Instead of letting things continue for two weeks as planned, Philip Zimbardo and his colleagues terminated the experiment after six days to avoid harming the subjects. Feelings of stress among the participants who played the role of prisoner seemed to be relieved by discussions in special postexperiment encounter sessions while follow-up during the next year indicated no lasting negative effects on the participants and some gains in self-insight.

FIGURE 3.4 ■ Chart of Guard and Prisoner Behavior



Source: Adapted from *The Lucifer Effect: Understanding How Good People Turn Evil by Philip G. Zimbardo*. Copyright 2007 by Philip G. Zimbardo, Inc. Used by permission of Random House, Inc., an imprint and division of Penguin Random House LLC; and Random House Group Ltd. All rights reserved.

Would you ban such experiments because of the potential for harm to subjects? Does it make any difference to you that Zimbardo's and Milgram's experiments seemed to yield significant insights into the effect of a social situation on human behavior—insights that could be used to improve prisons or perhaps lessen the likelihood of another holocaust (Reynolds 1979:133–139)? Do you believe this benefit outweighed the foreseeable risks? Arthur Miller (1986) argued that real harm "could result from *not doing* research on destructive obedience" (p. 138) and other troubling human behaviors. What if the researchers themselves—as was true of both Milgram (1974:27–31) and Zimbardo (2007)—did not foresee the potential harm to their participants?

Researchers who study specific communities, groups, or formal organizations should also consider how the reports and presentations they make after a research project concludes may affect the welfare of the collectivity and specific individuals within it. Will the collective's identity be obscured with a fictitious name? Would it be possible for members of the collective to identify individual participants if they read research reports? In the online social world of the twenty-first century, we cannot ensure even formal research publications will have a

limited audience. It is important to plan procedures for coping with unplanned publicity and for discussing the possibilities with members of the collective before findings are released (Lahman 2018:73).

Obtain Informed Consent

To be informed, consent must be given by the persons who are competent to consent, have consented voluntarily, are fully informed about the research and know who is conducting the research, and have comprehended what they have been told (Reynolds 1979). There was no expectation of obtaining informed consent from research participants when the Tuskegee Study began in the 1930s, and the researchers did their best to keep the men enrolled in the study unaware of its purpose for fear that the men would decline to participate (Benedek and Erlen 1999; Washington 2006).

Obtaining informed consent creates additional challenges for researchers. The researcher's actions and body language should help convey their verbal assurance that consent is voluntary. The language of the consent form must be clear and understandable to the research participants and yet sufficiently long and detailed to explain what will actually happen in the research. Using clear and simple language is essential (Lahman 2018:74). You probably also realize, as Baumrind (1985) did, that because of the inability to communicate perfectly, "full disclosure of everything that could possibly affect a given subject's decision to participate is not possible, and therefore cannot be ethically required" (p. 165). I present in the next section two examples of informed consent forms that represent different approaches.

As in Milgram's study, experimental researchers whose research design requires some type of subject deception try to get around the problem of providing fully informed consent by withholding some information before the experiment begins but then debriefing subjects at the end. In a **debriefing**, the researcher explains to the subjects what happened in the experiment and why and then responds to their questions. A carefully designed debriefing procedure can help the research participants learn from the experimental research and grapple constructively with feelings elicited by the realization that they were deceived (Sieber 1992:39–41). However, even though debriefing can be viewed as a substitute, in some cases, for securing fully informed consent before the experiment, debriefed subjects who disclose the nature of the experiment to other participants can contaminate subsequent results (Adair, Dushenko, and Lindsay 1985). Apparently for this reason, Milgram provided little information in his "debriefing" to participants in most of his experiments. It was only in the last two months of his study that he began to provide more information while still asking participants not to reveal the true nature of the experimental procedures until after the study was completely over (Perry 2013:76, 84). Unfortunately, if the debriefing process is delayed, the ability to lessen any harm resulting from the deception is also reduced.

For a study of the social background of men who engage in homosexual behavior in public facilities, Laud Humphreys (1970) decided that truly informed consent would be impossible to obtain. Instead, he first served as a lookout—a "watch queen"—for men who were entering a public bathroom in a city park with the intention of having sex. In a number of cases, he then left the bathroom and copied the license plate numbers of the cars driven by the men. One year later, he visited the homes of the men and interviewed them as part of a larger study of social issues. Humphreys changed his appearance so the men did not recognize him. In ***Tearoom Trade***, his book on this research, Humphreys concluded the men who engaged in what were then viewed as deviant acts were, for the most part, married, suburban men whose families were unaware of their sexual practices. But debate has continued ever since about Humphreys's failure

to tell the men what he was really doing in the bathroom or why he had come to their homes for the interview. He was criticized by many, including some faculty members at the University of Washington, who urged that his doctoral degree be withheld. However, many other professors and some members of the gay community praised Humphreys for helping normalize conceptions of homosexuality (Miller 1986:135).

If you were to serve on your university's IRB, would you allow this research to be conducted? Can students who are asked to participate in research by their professor be considered able to give informed consent? Do you consider *informed consent* to be meaningful if the true purpose or nature of an experimental manipulation is not revealed?

The process and even possibility of obtaining informed consent must consider the capacity of prospective participants to give informed consent. Children cannot legally give consent to participate in research; instead, they must in most circumstances be given the opportunity to give or withhold their *assent* to participate in research, usually by a verbal response to an explanation of the research. In addition, a child's legal guardian must give written informed consent to have the child participate in research (Sieber 1992). There are also special protections for other populations who are likely to be vulnerable to coercion or undue influence—defined in the new regulations as incarcerated persons, individuals with impaired decision-making ability, and educationally or economically disadvantaged persons (Chadwick 2017:4; *Federal Register*, 2017, p. 7263). Would you allow research on incarcerated persons, whose ability to give informed consent can be questioned? What special protections do you think would be appropriate? Are you reassured that the regulations allow research on incarcerated persons that is designed to evaluate or improve their conditions (Lahman 2018:18)?

Obtaining informed consent also becomes more challenging in some cultures or in collectivist communities in which leaders or the whole group are accustomed to making decisions for individual members (Lahman 2018). In such settings, usually in non-Western cultures, researchers may have to develop a relationship with the community before individuals can be engaged in research (Bledsoe and Hopson 2009:397–398). Regulations adopted by the United Nations in 2007 stipulate the right of indigenous communities to control access to research in their communities (Israel 2014:92).

Subject payments create yet another complication for achieving the goal of informed consent. Although payments to research participants can be a reasonable way to compensate them for their time and effort, payments also serve as an inducement to participate. If the payment is a significant amount in relation to the participants' usual income, it could lead people to set aside their reservations about participating in a project—even though they may harbor those reservations (Fisher and Anushko 2008:104–105).

Avoid Deception in Research, Except in Limited Circumstances

Deception occurs when subjects are misled about research procedures to determine how they would react to the treatment if they were not research subjects. The Tuskegee researchers' deceitful recruitment of research participants by pretending they were being enrolled in a free treatment program is an entirely different, blatant form of deception that could not be justified in any circumstances.

Deception about research procedures is a critical component of many social psychology experiments, partly because of the difficulty of simulating real-world stresses and dilemmas in a laboratory setting. The goal is to get subjects "to accept as true what is false or to give a false impression" (Korn 1997:4). In Milgram's (1963) experiment, for example, deception seemed necessary because the subjects could not be permitted to administer real electric shocks to the "stooge," yet it

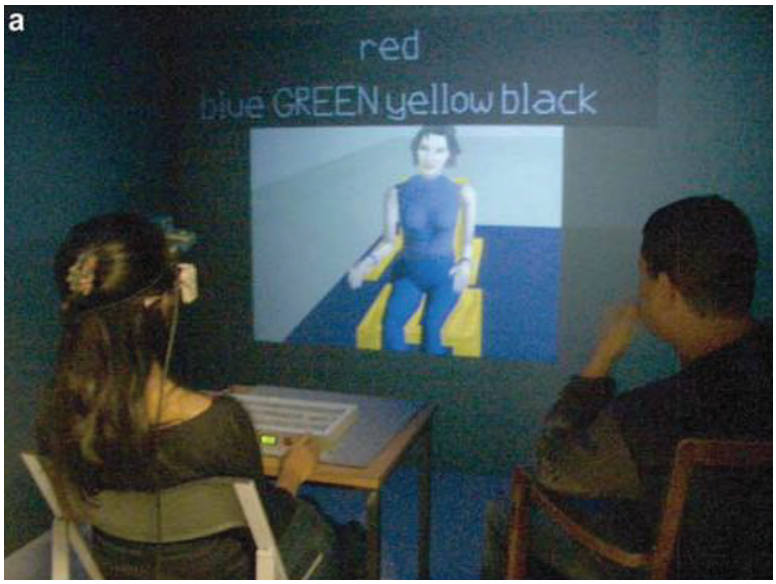
would not have made sense to order the subjects to do something they didn't find to be troubling. Milgram (1992:187–188) insisted the deception was absolutely essential, although the experimental records indicate that some participants figured out the deception (Perry 2013:128–129). The real question is Is this sufficient justification to allow the use of deception?

Gary Marshall and Philip Zimbardo (1979:971–972) sought to determine the physiological basis of emotion by injecting student volunteers with adrenaline, so their heart rates and sweating would increase, and then placing them in a room with a student "stooge" who acted silly. The students were told they were being injected with a vitamin supplement to test its effect on visual acuity (Korn 1997:2–3). Jane Allyn Piliavin and Irving Piliavin (1972:355–356) staged fake seizures on subway trains to study helpfulness (Korn 1997:3–4). George Schreer, Saundra Smith, and Kirsten Thomas (2009) investigated racial profiling by sending "customers" to browse in high-end retail stores and then observing the behaviors of salespersons. If you were a member of your university's IRB, would you vote to allow some or all such deceptive practices in research? What about less dramatic instances of deception in laboratory experiments with students like yourself?

The development of computer-based techniques for creating virtual reality environments provides a way to lessen this dilemma. For example, Mel Slater, Angus Antley, and a team of European researchers (2006) repeated the Milgram obedience experiment procedures with virtual reality techniques. According to participants' statements, behaviors, and physiological responses, they seemed to experience the immersive virtual environment as if it was real, even though they clearly understood the "learner" they observed was only virtual (see Figure 3.5). The participants also responded to the "experimenter's" requests in a way that was similar to what occurred in Milgram's "real" experiment.

The new federal regulations relax concerns about deception of the type often used in laboratory experiments in social psychology. Specifically, the regulations authorize deception in

FIGURE 3.5 ■ Virtual Reprise of Milgram's Obedience Experiments



Source: Slater M, Antley A, Davison A, Swapp D, Guger C. 2006. "A Virtual Reprise of the Stanley Milgram Obedience Experiments." *PLoS ONE* 1(1).

research "where the subject is informed they will be unaware of or misled regarding the nature or purposes of the research" (Chadwick 2017:4). Ethics codes in some European countries permit covert research (Israel 2014:99).

Maintain Privacy and Confidentiality

Maintaining privacy and confidentiality is another key ethical standard for protecting research participants, and the researcher's commitment to that standard should be included in the informed consent agreement (Sieber 1992). Procedures to protect each subject's privacy—such as locking records and creating special identifying codes—must be established to minimize the risk of access by unauthorized persons. However, statements about confidentiality should be realistic: Laws allow research records to be subpoenaed and may require reporting of child abuse; a researcher may feel compelled to release information if a health- or life-threatening situation arises and participants need to be alerted. It can be difficult to maintain confidentiality when research is reported about a specific community and involves particular behaviors or events that community members may recognize (Israel 2014:107–113). Also, the standard of confidentiality does not apply to observation in public places and information available in public records.

There is one exception to some of these constraints: The National Institutes of Health (NIH) can issue a **Certificate of Confidentiality** to protect researchers from being legally required to disclose confidential information. This is intended to help researchers overcome the reluctance of individuals engaged in illegal behavior to sign a consent form or to risk exposure of their illegal activities (Lahman 2018:81; Sharma 2009:426). Researchers who are focusing on high-risk populations or behaviors, such as crime, substance abuse, sexual activity, or genetic information, can request such a certificate. Suspicions of child abuse or neglect must still be reported, and in some states, researchers may still be required to report such crimes as elder abuse (Arwood and Panicker 2007).

The **Health Insurance Portability and Accountability Act (HIPAA)** passed by Congress in 1996 created more stringent regulations for the protection of health care data. As implemented by the DHHS in 2000 (revised in 2002), the HIPAA Final Privacy Rule applies to oral, written, and electronic information that "relates to the past, present or future physical or mental health or condition of an individual." The HIPAA rule requires that researchers have valid authorization for any use or disclosure of "protected health information" from a health care provider. Waivers of authorization can be granted in special circumstances (Cava, Cushman, and Goodman 2007).

Consider Uses of Research So That Benefits Outweigh Risks

Scientists must also consider the uses to which their research is put. Although many scientists believe that personal values should be left outside the laboratory, some feel it is proper—even necessary— for scientists to concern themselves with the way their research is used.

Milgram made it clear he was concerned about the phenomenon of obedience precisely because of its implications for people's welfare. As you have already learned, his first article (Milgram 1963) highlighted the atrocities committed under the Nazis by citizens and soldiers who were "just following orders." In his more comprehensive book on the obedience experiments (Milgram 1974), he also argued that his findings shed light on the atrocities committed in the Vietnam War at My Lai, slavery, the destruction of the Native American population, and the internment of Japanese Americans during World War II. Milgram makes no explicit attempt to "tell us what to do" about this problem. In fact, as a dispassionate social scientist, Milgram (1974) says, "What the present study [did was] to give the dilemma [of obedience to authority]

contemporary form by treating it as subject matter for experimental inquiry, and with the aim of understanding rather than judging it from a moral standpoint" (p. xi).

Yet it is impossible to ignore the very practical implications of Milgram's investigations, which Milgram took pains to emphasize. His research highlighted the extent of obedience to authority. Although it was not as widely discussed, Milgram's various versions of the obedience experiment also identified multiple factors that could be manipulated to lessen blind obedience, including encouraging dissent by at least one group member, removing the subject from direct contact with the authority figure, and increasing the contact between the subject and the victim. It is less clear how much Milgram's laboratory manipulation can tell us about obedience in the very different historical events to which he generalized his conclusions, but some subsequent research tried to evaluate this. In the Hofling Hospital Experiment, real nurses obediently administered a lethal dose of a (fake) medicine when requested by a (fake) doctor on the phone (Hofling et al. 1966)—although this was not true when the command was given in person by a real doctor about a real medicine and the nurses had time to consult each other (Rank and Jacobson 1977). Do you agree that this type of research can have potentially great benefits for society?

Social scientists who conduct research on behalf of specific organizations may face additional difficulties when the organization, instead of the researcher, controls the final report and the publicity it receives. If organizational leaders decide that particular research results are unwelcome, the researcher's desire to have the findings used appropriately and reported fully can conflict with contractual obligations. Researchers can often anticipate such dilemmas in advance and resolve them when the contract for research is negotiated—or they may simply decline a particular research opportunity altogether. But often, such problems come up only after a report has been drafted, or the problems are ignored by a researcher who needs to have a job or needs to maintain particular personal relationships. These possibilities cannot be avoided entirely, but because of them, it is always important to acknowledge the source of research funding in reports and to consider carefully the sources of funding for research reports written by others.

The potential of withholding a beneficial treatment from some subjects also is a cause for ethical concern. One of the strongest charges of ethical violations in The Tuskegee Study is about the decision not to offer penicillin treatment to the research participants, even after it was widely accepted as effective for syphilis (Jones 1993). Even in this case, however, some argue that knowledge was limited at the time about the benefits of penicillin in late stages of the disease (Benedek and Erlen 1999:25). In some projects, the safety of the researchers and their staff also requires consideration. Although it doesn't happen often, research in some social settings about organized crime and in some countries about insurgent movements and military abuses have resulted in researchers being killed, injured, or arrested. Interviewers or observers may also suffer from emotional reactions to harm done to others, including those they interview. No research project should begin without careful consideration of whether researcher and staff safety may be an issue. If it is, a safety protocol must be developed in which procedures for ensuring safety are spelled out (Fawcett and Pockett 2015:138–139; Israel 2014:174–179).

We also need to consider the impact of research on the prospects for subsequent research projects. The ethical violations of the Tuskegee Study contributed to mistrust in the Black community of the medical profession and of medical research (Hoberman 2012). This in turn has contributed to less participation in medical research (Scharff et al. 2010) and to more vaccine hesitancy during the COVID-19 pandemic (Callaghan et al. 2021; Cohen and Vigue 2020; Ferdinand 2021). These effects in turn increase risks of illness and death and slow the progress of research that can improve outcomes (Jacobs 2021).

RESEARCH IN THE NEWS
A HOLISTIC VIEW OF VACCINE HESITANCY

New York Times opinion columnist Charles M. Blow contrasted his orientation to getting the COVID-19 vaccine with that of a younger male friend as "Absolutely" versus "Absolutely not." Wondering whether this difference between two Black men might generalize to the Black population, he found age variations reported in a survey by the National Foundation for Infectious Diseases. Among Black adults 60 and older, 68% planned to get the vaccine, compared to only 38% of those 18 to 44. Mr. Blow found this discrepancy hard to explain in relation to the effects of the Tuskegee experiments on trust in the medical system, since the younger Black people who were most hesitant to get the vaccine were not even alive when the experiment was conducted or exposed. Instead, he speculated, the explanation is more mistrust of government among younger people who came of age during a period of mass incarceration, widely publicized police killings of Black people, and inadequate responses to disasters like Hurricane Katrina and the Flint water crisis.

For Further Thought?

1. What events have influenced your level of trust in the government? Do you think your own orientation to the vaccine has been affected by this level of trust?
2. What are some steps you would recommend to improve the acceptance of social research projects among members of disadvantaged communities?

News source: Blow, Charles M. 2021. "A Holistic View of Vaccine Hesitancy." *The New York Times*, February 7.

INFORMED CONSENT FORMS

Consent forms A (Figure 3.6) and B (Figure 3.7) illustrate two different approaches to the trade-off in obtaining informed consent between providing clear and understandable explanations and sufficient details about research procedures.

Consent Form A was approved by my university IRB for in-person interviews with current and former applicants to and participants in a community program for adolescents. It is brief and to the point.

Consent Form B reflects the requirements of an academic hospital's IRB (I have included only a portion of the six-page form). Because the hospital is used to reviewing research proposals involving drugs and other treatment interventions with hospital patients, it requires a very detailed and lengthy explanation of procedures and related issues, even for a simple interview study such as mine with Dr. Schapira. You can probably imagine the requirement that prospective participants sign such lengthy consent forms can reduce their willingness to participate in research and perhaps influences their responses if they do agree to participate (Larson 1993:114).

A lengthy consent form can also be difficult for participants to understand when it discusses issues like medical treatments and possible side effects. The new federal human subjects regulations adopted in 2017 include a requirement for an initial concise explanation in consent forms that is intended to improve participant understanding (*Federal Register* 2017:7265):

> Informed consent must begin with a concise and focused presentation of the key information that is most likely to assist a prospective subject or legally authorized representative in understanding the reasons why one might or might not want to participate in the research. This part of the informed consent must be organized and presented in a way that facilitates comprehension.

FIGURE 3.6 ■ Consent Form A

University of Massachusetts Boston
Department of Sociology
100 Morrissey Boulevard
Boston, MA 02125-3393

Consent Form for Teen Empowerment Study: Current Organizers

Introduction and Contact Information

I am conducting a study of the Teen Empowerment Program. In order for you to participate, I need to ask for your consent.

The principal researcher is Professor Russell Schutt, Department of Sociology at UMass Boston. I am Whitney Gecker, his research assistant in this project. Please read this form and feel free to ask questions. If you have further questions later, Professor Schutt will discuss them with you. His telephone number is 617-287-6253.

Description of the Project

Teen Empowerment (TE) is an innovative program for youth in Somerville, Massachusetts and several other locations. This research project is a comprehensive evaluation of the Teen Empowerment program in Somerville, MA. It is to learn whether and if so how the TE model is successful. The study will describe the backgrounds, attitudes, experiences and activities of current and past program participants and of youth who applied to the program but did not participate in it. The study will identify the impact of the TE process on individual participants, including their behavior, their feelings about themselves, and their awareness and demonstration of love and forgiveness toward others. It will also examine the effect of TE on participants' relationships with others; on the rate of crime in the community, and on the participants' lives in comparison to the lives of program applicants who were not accepted.

The research procedures involve interviews with current and former youth participants and unsuccessful applicants, as well as observation of group activities. Participation in this study will take approximately one hour. If you decide to participate in this study, you will be asked a series of questions about your attitudes, your experiences in school and work, your relations with family and friends, your health behaviors and experiences of violence. You will also be asked about the Teen Empowerment program and the Somerville community. You will receive a $15 gift certificate for your participation. If you agree, we will also record information that you provided to Teen Empowerment on forms you collected when you applied to the program.

Risks or Discomforts

The primary risk associated with this study is that the investigators may learn information about you that you did not wish to share with them. If you have any concern that you have disclosed information that you do not wish to have known, or if you have any other concerns or feelings as a result of the interview, you may speak with Professor Schutt to discuss this at any point, by calling 617-287-6253, or by communicating by email (Russell.schutt@umb.edu) or regular mail, addressed to Professor Schutt at the University of Massachusetts, 100 Morrissey Blvd., Boston, MA 02125-3393. If you indicate any feelings of distress during the study, the project interviewer will suggest that you speak to a Teen Empowerment staff member and will provide referral information for service agencies in Somerville.

Confidentiality

Your part in this research is confidential. The final project report will not name the specific individuals who were interviewed nor link comments made in interviews to these individuals. Please note that Massachusetts law requires that information about abuse or neglect must be reported to the Massachusetts Department of Children and Families. If you state that you are abused or neglected, I will inform Professor Schutt, the Principal Investigator, and he will report this information to the Massachusetts Department of Children and Families.

Voluntary Participation

The decision whether or not to take part in this research study is voluntary. If you do decide to take part in this study, you may decline to answer any question and you may terminate participation at any time without consequence. Whatever you decide about participation will not be known to any other current or former Teen Empowerment participants or staff or to anyone else.

Rights

You have the right to ask questions about this research at any time during the study. You can reach Russell Schutt to ask further questions at 617-287-6253. If you have any questions or concerns about your rights as a research participant, you may contact a representative of the Institutional Review Board (IRB), at the University of Massachusetts, Boston, which oversees research involving human participants. The Institutional Review Board may be reached at the following address: IRB, Quinn Administration Building-2-015, University of Massachusetts Boston, 100 Morrissey Boulevard, Boston, MA 02125-3393.You can also contact the Board by telephone or e-mail at (617) 287-5370 or at human.subjects@umb.edu.

I HAVE READ THE CONSENT FORM. MY QUESTIONS HAVE BEEN ANSWERED. MY SIGNATURE ON THIS FORM INDICATES THAT I CONSENT TO PARTICIPATE IN THIS STUDY. I ALSO CERTIFY THAT I AM 18 YEARS OF AGE OR OLDER.

_____ _____

Signature of Participant Date

_____

Typed/Printed Name of Participant

MY SIGNATURE BELOW INDICATES THAT I CONSENT TO HAVE INFORMATION USED IN THE STUDY THAT I PROVIDED WHEN I APPLIED TO TEEN EMPOWERMENT.

_____ _____

Signature of Participant Date

_____ _____

Signature of Researcher Date

_____

Typed/Printed Name of Researcher

FIGURE 3.7 ■ Consent Form B

Research Consent Form for Social and Behavioral Research

Dana-Farber/Harvard Cancer Center

BIDMC/BWH/CH/DFCI/MGH/Partners Network Affiliates OPRS 11-05

Protocol Title: ASSESSING COMMUNITY HEALTH WORKERS' ATTITUDES AND KNOWLEDGE ABOUT EDUCATING COMMUNITIES ABOUT CANCER CLINICAL TRIALS

DF/HCC Principal Research Investigator / Institution: Dr. Russell Schutt, PhD / Beth Israel Deaconess Medical Center and Univ. of Massachusetts, Boston

DF/HCC Site-Responsible Research Investigator(s) / Institution(s): Lidia Schapira, MD /Massachusetts General Hospital

Interview Consent Form

A. INTRODUCTION

We are inviting you to take part in a research study. Research is a way of gaining new knowledge. A person who participates in a research study is called a "subject." This research study is evaluating whether community health workers might be willing and able to educate communities about the pros and cons of participating in research studies.

It is expected that about 10 people will take part in this research study.

An institution that is supporting a research study either by giving money or supplying something that is important for the research is called the "sponsor." The sponsor of this protocol is National Cancer Institute and is providing money for the research study.

This research consent form explains why this research study is being done, what is involved in participating in the research study, the possible risks and benefits of the research study, alternatives to participation, and your rights as a research subject.The decision to participate is yours. If you decide to participate, please sign and date at the end of the form. We will give you a copy so that you can refer to it while you are involved in this research study.

If you decide to participate in this research study, certain questions will be asked of you to see if you are eligible to be in the research study. The research study has certain requirements that must be met. If the questions show that you can be in the research study, you will be able to answer the interview questions.

If the questions show that you cannot be in the research study, you will not be able to participate in this research study.

DFCI Protocol Number: 06-085	Date DFCI IRB Approved this Consent Form: January 16, 2007
Date Posted for Use: January 16, 2007	Date DFCI IRB Approval Expires: August 13, 2007

We encourage you to take some time to think this over and to discuss it with other people and to ask questions now and at any time in the future.

B. WHY IS THIS RESEARCH STUDY BEING DONE?

Deaths from cancer in general and for some specific cancers are higher for black people compared to white people, for poor persons compared to nonpoor persons, and for rural residents compared to non-rural residents. There are many reasons for higher death rates between different subpopulations. One important area for changing this is to have more persons from minority groups participate in research about cancer. The process of enrolling minority populations into clinical trials is difficult and does not generally address the needs of their communities. One potential way to increase participation in research is to use community health workers to help educate communities about research and about how to make sure that researchers are ethical. We want to know whether community health workers think this is a good strategy and how to best carry it out.

C.WHAT OTHER OPTIONS ARE THERE?

Taking part in this research study is voluntary. Instead of being in this research study, you have the following option:

Decide not to participate in this research study.

Research Consent Form for Social and Behavioral Research

Dana-Farber/Harvard Cancer Center

BIDMC/BWH/CH/DFCI/MGH/Partners Network Affiliates OPRS 11-05

D.WHAT IS INVOLVED IN THE RESEARCH STUDY?

Before the research starts (screening): After signing this consent form, you will be asked to answer some questions about where you work and the type of communityhealth work you do to find out if you can be in the research study.

If the answers show that you are eligible to participate in the research study, you will be eligible to participate in the research study. If you do not meet the eligibility criteria, you will not be able to participate in this research study.

After the screening procedures confirm that you are eligible to participate in the research study: You will participate in an interview by answering questions from a questionnaire. The interview will take about 90 minutes. If there are questions you prefer not to answer we can skip those questions. The questions are about the type of work you do and your opinions about participating in research. If you agree, the interview will be taped and then transcribed. Your name and no other information about you will be associated with the tape or the transcript. Only the research team will be able to listen to the tapes.

DFCI Protocol Number: 06-085	Date DFCI IRB Approved this Consent Form: January 16, 2007
Date Posted for Use: January 16, 2007	Date DFCI IRB Approval Expires: August 13, 2007

Immediately following the interview, you will have the opportunity to have the tape erased if you wish to withdraw your consent totaping or participation in this study. You will receive $30.00 for completing this interview.

After the interview is completed: Once you finish the interview there are no additional interventions.

. . .

N. DOCUMENTATION OF CONSENT

My signature below indicates my willingness to participate in this research study and my understanding that I can withdraw at any time.

_____ _____

Signature of Subject or Legally Authorized Representative — Date

_____ _____

Person obtaining consent — Date

To be completed by person obtaining consent:

The consent discussion was initiated on _____ (date) at _____ (time.)

☐ A copy of this signed consent form was given to the subject or legally authorized representative.

For Adult Subjects

☐ The subject is an adult and provided consent to participate.

☐ The subject is an adult who lacks capacity to provide consent and his/her legally authorized representative:

- ☐ gave permission for the adult subject to participate
- ☐ did not give permission for the adult subject to participate

DFCI Protocol Number: 06-085	Date DFCI IRB Approved this Consent Form: January 16, 2007
Date Posted for Use: January 16, 2007	Date DFCI IRB Approval Expires: August 13, 2007

Key information is to include identification of the consent as involving voluntary participation in research, its purposes and expected duration and procedures, and reasonably foreseeable risks and benefits, as well as alternative treatments (Chadwick 2017:5).

THE INSTITUTIONAL REVIEW BOARD

Federal regulations now require that every institution seeking federal funding for biomedical or behavioral research on human subjects have an **institutional review board** that reviews research proposals involving human subjects—including data about living individuals. According to federal regulations [45 CFR 46.102(d)], research is "a systematic investigation . . . designed to develop or contribute to generalizable knowledge," and according to the DHHS [45 CFR 46.102 (f)], a human subject is "a living individual about whom an investigator (whether professional or student) conducting research obtains data through intervention or interaction with the individual or just identifiable private information." The IRB determines whether a planned activity is research or involves human subjects.

IRBs at universities and other agencies apply ethical standards that are set by federal regulations but can be expanded or specified by the institution's IRB and involve all research at the institution irrespective of the funding source (Sieber 1992:5, 10). The IRB has the authority to require changes in a research protocol or to refuse to approve a research protocol if it deems human subjects protections inadequate. Consent forms must include contact information for the IRB, and the IRB has the authority to terminate a research project that violates the procedures the IRB approved or that otherwise creates risks for human subjects. The **Office for Human Research Protections, National Institutes of Health** monitors IRBs, with the exception of research involving drugs (which is the responsibility of the U.S. Food and Drug Administration).

To promote adequate review of ethical issues, the regulations require that IRBs include at least five members, with at least one nonscientist and one from outside the institution (Speiglman and Spear 2009:124). The new proposed regulations for IRBs also stipulate that they "shall be sufficiently qualified through the experience and expertise of its members, including race, gender, and cultural backgrounds and sensitivity to such issues as community attitudes, to promote respect for its advice and counsel in safeguarding the rights and welfare of human subjects (*Federal Register* 2017: 7263). When research is reviewed concerning populations vulnerable to coercion or undue influence, such as incarcerated persons, the IRB must include a member knowledgeable about that population (Chadwick 2017:4). Sensitivity to community attitudes and training in human subjects protection procedures is also required (Selwitz, Epley, and Erickson 2013).

Every member of an institution with an IRB—including faculty, students, and staff at a college or university—must submit a proposal to their IRB before conducting research with identifiable people. The IRB proposal must include research instruments and consent forms, as applicable, as well as enough detail about the research design to convince the IRB members that the potential benefits of the research outweigh any risks (Speiglman and Spear 2009:124). Most IRBs also require that researchers complete a training program about human subjects, usually the Collaborative Institutional Training Initiative (CITI) (https://about.citiprogram.org/en/homepage/). CITI training is divided into topical modules ranging from history, ethical principles, and informed consent to vulnerable populations, Internet-based research, educational research, and records-based research. Each IRB determines which CITI training modules researchers at its institution must complete.

Although the IRB is the responsible authority within the institution (Hicks 2013), in the new regulations, many proposals developed by social scientists are exempt from review because they involve very low perceived risk. These exemptions include (1) research about educational procedures in an educational setting, (2) "educational tests, survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording)," or (3) "research involving benign behavioral interventions in conjunction with the collection of information from an adult subject through verbal or written responses . . . or audiovisual recording" when the subject had agreed to the intervention (*Federal Register* 2017:7261, 7262, 7264). Criteria (2) and (3) also include the stipulation that at least one of the following criteria is met:

- The identity of the human subjects *cannot* readily be ascertained.
- Disclosure of the human subjects' responses would not place them at risk of criminal or civil liability or damage their financial standing, employability, educational advancement, or reputation.
- The identity of the human subjects *can* readily be ascertained, but a limited IRB review determines that procedures are in place that will protect the privacy of subjects and maintain the confidentiality of data.

Many research projects must be reviewed before the full IRB (Speiglman and Spear 2009:125–126). An IRB must ensure that several specific standards are met by research projects that it reviews either on an expedited basis or in a full board review (Hicks 2013):

1. Risks to subjects are minimized: (i) By using procedures that are consistent with sound research design and that do not unnecessarily expose subjects to risk, and (ii) whenever appropriate, by using procedures already being performed on the subjects for diagnostic or treatment purposes.
2. Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result. In evaluating risks and benefits, the IRB should consider only those risks and benefits that may result from the research (as distinguished from risks and benefits of therapies subjects would receive even if not participating in the research). The IRB should not consider possible long-range effects of applying knowledge gained in the research (for example, the possible effects of the research on public policy) as among those research risks or benefits that fall within the purview of its responsibility.
3. Selection of subjects is equitable. In making this assessment, the IRB should consider the purposes of the research and the setting in which the research will be conducted and should be particularly cognizant of the special problems of research involving vulnerable populations.

Informed consent is also required and, when appropriate, a provision for monitoring data collection to ensure subject safety. When some or all of the research participants are likely to be vulnerable to coercion or undue influence, additional safeguards have been included in the study to protect the rights and welfare of these subjects.

In addition, the IRB may serve as the privacy board that ensures researchers' compliance with HIPAA. In this capacity, the IRB responds to requests for waivers or alterations of the authorization requirement under the privacy rule for uses and disclosures of protected health

information in research. Researchers seeking to collect or use existing HIPAA data must provide additional information to the IRB about their plans for using the health information.

Ethical research practice does not end with formal IRB approval of procedures to protect human subjects and adherence to those procedures throughout the research process. Unanticipated ethical issues may arise after a study begins, and it is important to have a plan in place to consult with others and to check in with the IRB to develop appropriate responses. Many ethicists now emphasize the importance of ongoing evaluation and response as ethical issues may arise during a research project (Lahman 2018; Poth 2021).Consider adopting the approach of **process responsiveness** by checking with participants during a study about their comfort with the study activities and their continued consent to participate (Lahman 2018).

SOCIAL RESEARCH PROPOSALS

Now that you have an overview of the research process and a basic understanding of IRB requirements, it is time to introduce the process of writing a research proposal. A research proposal is the launching pad for a formal research project, and it serves the very important function of forcing you, as the researcher, to set out a problem statement and a research plan—to think through the details of what you are trying to accomplish and how you will go about that—as well as to think through procedures for the protection of human subjects. So whether you must write a proposal for a professor, a thesis committee, an organization seeking practical advice, or a government agency that funds basic research, you should approach the requirement as a key step toward achieving your goals. Just writing down your ideas will help you see how they can be improved, and almost any feedback will help you refine your plans.

Each chapter in this book includes a section (online at the study site), "Developing a Research Proposal," with exercises that guide you through the process of proposal writing. This section introduces the process of proposal writing. It also provides a schematic overview of the entire research process. You will want to return to this section frequently so you will remember "where you are" in the research process as you learn about particular methods in the remaining chapters.

Research proposals often have five sections (Locke, Spirduso, and Silverman 2000:8–34):

- *An introductory statement of the research problem*, in which you clarify what it is you are interested in studying
- *A literature review*, in which you explain how your problems and plans build on what has already been reported in the literature on this topic
- *A methodological plan*, detailing just how you will respond to the particular mix of opportunities and constraints you face
- *An ethics statement*, identifying human subjects issues in the research and how you will respond to them in an ethical fashion
- *A statement of limitations*, reviewing the potential weaknesses of the proposed research design and presenting plans for minimizing their consequences

You will also need to include a budget and project timeline, unless you are working within the framework of a class project.

When you develop a research proposal, it will help to use a checklist like that in Figure 3.8; see also Gregory Herek (1995). It is easy to omit important details and to avoid being self-critical while rushing to put together a proposal. However, it is even more painful to have a proposal

FIGURE 3.8 ■ Decisions in Research

Problem Formulation (Chapters 1–3)

1. Developing a research question
2. Assessing researchability of the problem
3. Consulting prior research
4. Relating to social theory
5. Choosing an approach: Deductive? Inductive? Descriptive?
6. Reviewing research guidelines and ethical standards

Checkpoint 1
Alternatives:

- Continue as planned.
- Modify the plan.
- Stop. Abandon the plan.

Research Validity (Chapters 4–6)

7. Establishing measurement validity:
   - How should concepts be defined?
   - What measures are available, and what measures must be developed?
   - How can reliability and validity be assessed?
   - Is authenticity an important goal, and how can it be assessed?
8. Establishing generalizability:
   - Was a representative sample used?
   - Are the findings applicable to particular subgroups?
   - Does the population sampled correspond to the population of interest?
9. Establishing causality:
   - What is the possibility of experimental or statistical controls?
   - How can the causal mechanism be assessed?
   - What is the causal context?
10. Determining the data required: Longitudinal or cross-sectional?
11. Determining the units of analysis: Individuals or groups?
12. Determining the major possible sources of causal invalidity

Checkpoint 2
Alternatives:

- Continue as planned.
- Modify the plan.
- Stop. Abandon the plan.

Research Design (Chapters 7, 8, 10, 12–15)

13. Choosing a research design and procedures: Experimental? Survey? Participant observation? Historical, comparative? Evaluation research? Secondary data analysis? Mixed methods?
14. Specifying the research plan: Type of surveys, observations, etc.
15. Determining secondary analysis and availability of suitable data sets
16. Choosing a causal approach: Idiographic or nomothetic?
17. Assessing human subjects protections

Checkpoint 3
Alternatives:

- Continue as planned.
- Modify the plan.
- Stop. Abandon the plan.

(*Continued*)

FIGURE 3.8 ■ Decisions in Research (*Continued*)

Data Analysis (Chapters 9, 11)

18. Choosing an analytic approach:
    - Identifying statistics and graphs for describing data
    - Identifying relationships between variables
    - Deciding about statistical controls
    - Testing for interaction effects
    - Evaluating inferences from sample data to the population
    - Developing a qualitative analysis approach

Checkpoint 4
Alternatives:
- Continue as planned.
- Modify the plan.
- Stop. Abandon the plan.

Reporting Research (Chapter 16)

19. Clarifying research goals and prior research findings
20. Identifying the intended audience
21. Developing tables and charts
22. Organizing the text
23. Reviewing research limitations

Checkpoint 5
Alternatives:
- Complete the research!
- Modify the plan.
- Stop. Abandon the plan.

rejected (or to receive a low grade). It is better to make sure the proposal covers what it should and confronts the tough issues that reviewers (or your professor) will be sure to consider.

The list in Figure 3.8 can serve as a map to subsequent chapters in this book and as a checklist of decisions that must be made throughout any research project. The checklist is organized in five sections, each concluding with a checkpoint at which you should consider whether to proceed with the research as planned, modify the plans, or stop the project altogether. The sequential ordering of these steps obscures a bit the way in which they should be answered: not as single questions, one at a time, but as a unit—first as five separate stages and then as a whole. Feel free to change your checkpoint decisions based on your decisions about other points in the list.

We will learn how to apply the decision checklist with an example from a proposal focused on a public health care coordination program. At this early point in your study of research methods, you may not recognize all the terms in this checklist. Don't let that bother you now because my goal is just to give you a quick overview of the decision-making process. Your knowledge of these terms and your understanding of the decisions will increase as you complete each chapter.

Case Study: Evaluating a Public Health Program

Figure 3.9 provides excerpts from the research proposal I submitted to our IRB as part of an evaluation of a public health program for low-income residents funded by the U.S. Centers for Disease Control and the Massachusetts Department of Public Health (DPH) (Schutt 2011a). Appendixes included consent forms, research instruments, and the bibliography.

FIGURE 3.9 ■ An IRB Proposal for a Program Evaluation

Evaluation of the Coordinated Care Program

Rationale

Inadequate personal resources, knowledge, and service opportunities limit the access of low-income and minority populations to health care services. Percac-Lima and colleagues (2010) identified four types of barriers to effective service delivery in interviews with Latina women using a health care center: anxiety or fear and concerns about pain associated with the procedure ("personal barriers") and inadequate communication and scheduling difficulties ("system barriers"). Vargas et al. (2008) identified additional systemic barriers of provider bias and poor quality of care. Many public health agencies have adopted programs designed to overcome these barriers and thus to reduce disparities in health care services and, ultimately, outcomes.

The Massachusetts Department of Public Health began a Coordinated Care Program in 2006 in order to improve identification of health needs among low-income uninsured and underinsured residents and to increase the effectiveness of health services for this population. . . .

The purpose . . . is to investigate patient experience with the program, including subjective outcomes of the program and the bases of these outcomes. . . .

Program Background

Congress funded the National Breast and Cervical Cancer Early Detection Program (Breast and Cervical Cancer Mortality Prevention Act of 1990) to improve the rate of screening, testing, and referral to treatment for low-income uninsured and underinsured women at risk of breast and cervical cancer . . .

. . . . Programs offering patient navigation and care coordination have not typically been combined together due to their different disease foci: cancer for patient navigation and chronic illness for care coordination. The MA CCP program thus provides a unique opportunity to study the effect of combining these two different types of programs and examining their value for a very diverse and historically underserved population.

HIPAA Compliance Information

. . . No information about health conditions will be reported in a way that could be linked to any particular client.

Methodology

. . . Data will be collected with a statewide phone survey of a representative sample and through in-person interviews with a small number of clients at six of the 17 program delivery sites.
The phone survey is to result in a total sample of 400 interviews. One thousand clients will be sent an informational letter about the survey, indicating that they may receive a call. Cases will be selected randomly from the initial sampling frame and calledto determine study eligibility. Phone calls will be repeated up to 20 times in order to make contact. Questions S0-S4 (in theattached questionnaire) will serve as the screening questions. . . . The phone survey instrument will be translated into Spanish and Portuguese. . . .

Questions in the phone survey instrument are designed to identify relation to the health center and patient navigator, . . . perceived barriers to the use of health services, and satisfaction with health services. . . .

A sample of 30 clients will be selected for more intensive interviews about program experience and orientations. . . . For those clients who consent, the in-person interviews will be recorded and transcribed. The resulting textual data will be coded and analyzed with the assistance of qualitative analysis software . . .

Human Participants Information

Participant Data

The client database maintained by the Massachusetts Care Coordination Program identifies . . . active clients. . . .

Risks and Benefits

Participation in the Client Survey involves a minimal risk of loss of confidentiality. In addition, the probability and magnitude of harm or discomfort anticipated from responding to the Client Survey items is no greater than what would ordinarily be encountered in daily life, that is, minimal risk. Although the only direct benefit to survey participants is the opportunity to express their perceptions of and satisfaction with . . . case management, the information gained may be helpful to case managers and other health care providers. . . . The potential benefits of the proposed survey outweigh the minimal risk for individual survey participants.

(*Continued*)

FIGURE 3.9 ■ An IRB Proposal for a Program Evaluation (*Continued*)

Informed Consent and Confi dentiality

The . . . phone survey interviewer will inform the clients that the survey is voluntary and explain procedures to be followed (the respondent can stop at any time; can skip any question she does not want to answer). . . . . [P]hone interviewers and other staff adhere to strict confidentiality requirements. Identifying information will be used only for client contact and no personal identifiers will be maintained in the phone survey database. The phone survey ID numbers will be stored in a locked file . . .
All in-person interviews will be taped and transcribed and records will also be stored in a locked cabinet and on password-protected computers. . . . Consent forms will include consent for the interview, consent for taping, and consent for follow-up.

As you can see from the excerpts, I proposed to evaluate a care coordination program for low-income uninsured and underinsured Massachusetts residents (before universal health care). The proposal included a lengthy literature review, a description of the population and the sampling procedure, measures to be used in the survey, and the methods for conducting phone interviews as well as in-person interviews with a subset of the sample. Required sections for the IRB also included a statement of risks and benefits, procedures for obtaining informed consent, and means for maintaining confidentiality. A HIPAA compliance statement was included because of the collection of health-related data.

Let's review the issues identified in Figure 3.8 as they relate to the public health proposal. The research question concerned the effectiveness of a care coordination program involving the use of patient navigators—an evaluation research question [Question 1]. This problem certainly was suitable for social research, and it was one that was feasible with the money DPH had committed [2]. Prior research demonstrated clearly that the program had potential but also that this approach had not previously been studied [3]. The treatment approach was connected to theories about health care disparities [4] and, given prior work and uncertainties in this area, mixed methods involving both a deductive, hypothesis-testing approach and an inductive, exploratory approach were called for [5]. I argued to the IRB that our plan protected human subjects and took each research guideline into account [6]. So it seemed reasonable to continue to develop the proposal (Checkpoint 1).

Measures were to include structured survey questions to measure many variables and open-ended questions to allow exploration of barriers to care [7]. Use of a representative sample of the population of program recipients would increase the generalizability of the findings, although I was limited to interviews in English, Spanish, and Portuguese, even though we knew there were a small number of recipients who spoke other languages [8]. The problem was well suited to a survey design in which questions about causality could be addressed with statistical controls, with some attention to contextual differences between different service sites. However, there were to be no tests of causal hypotheses or investigation of causal mechanism since the program was already in place and there was no opportunity to design a randomized experiment of the effect of the new program [9]. Cross-sectional data [10], involving individuals [11], could be used to investigate reactions to the program. The design left several sources of causal invalidity, including the possibility that persons who received the most services from patient navigators were those who had more resources and so were more healthy [12]. It seemed that I would be able to meet basic criteria for validity, but there would be uncertainty about the causal effect of the new program features (Checkpoint 2).

A survey design was preferable because this was to be a study of a statewide population, but I did include a qualitative component so that I could explore orientations [13, 14]. Because the effectiveness of the program strategy had not been studied before in this type of population, I could not

propose doing a secondary data analysis or meta-analysis [15]. I sought only to investigate causation from a *nomothetic* perspective, without attempting to show how the particular experiences of each participant may have led to their outcome [16]. The study design was low risk and included voluntary participation; the research design seemed ethical [17] (Checkpoint 3). Standard statistical tests were proposed as well as some analysis of qualitative data [18] (Checkpoint 4). My goal was to use the research as the basis for several academic articles, as well as a report to the agency [19, 20]. Of course at this point, I was not ready to develop tables and charts [21] or organize the text of the report [22], but I was starting to think about study methodology's limitations [23] (Checkpoint 5).

If your research proposal will be reviewed competitively, it must present a compelling rationale for funding. You should not overstate the importance of the research problem that you propose to study (see the first section of this chapter). If you propose to test a hypothesis, be sure it is one for which there are plausible alternatives. You want to avoid focusing on a "boring hypothesis"—one that has no credible alternatives, even though it is likely to be correct (Dawes 1995:93).

A research proposal also can be strengthened considerably by presenting results from a pilot study of the research question. This might have involved administering the proposed questionnaire to a small sample, conducting a preliminary version of the proposed experiment with a group of students, or making observations over a limited period in a setting like that proposed for a qualitative study. My original proposal to the DPH was strengthened by my ability to build on a prior study I had conducted in the agency 10 years previously. Careful presentation of the methods used in the pilot study and the problems that were encountered will impress anyone who reviews the proposal.

Don't neglect the procedures for the protection of human subjects. Even before you begin to develop your proposal, you should find out what procedure your university's IRB requires for the review of student research proposals. Follow these procedures carefully, even if they require you to submit your proposal for an IRB review. No matter what your university's specific requirements are, if your research involves human subjects, you will need to include in your proposal a detailed statement that describes how you will adhere to these requirements.

By the book's end, in Chapter 16, you will have attained a much firmer grasp of the various research decisions outlined in Figure 3.8.

CONCLUSIONS

The extent to which ethical issues are a problem for researchers and their subjects varies dramatically with the type of research design. Survey research, in particular, often creates few ethical problems. In fact, researchers from Michigan's Institute for Survey Research interviewed a representative national sample of adults some years ago and found that 68% of those who had participated in a survey were somewhat or very interested in participating in another; the more times respondents had been interviewed, the more willing they were to participate again. Presumably, they would have felt differently if they had been treated unethically (Reynolds 1979:56–57). Conversely, some experimental studies in the social sciences that have put people in uncomfortable or embarrassing situations have generated vociferous complaints and years of debate about ethics (Reynolds 1979; Sjoberg 1967).

The evaluation of ethical issues in a research project should be based on a realistic assessment of the overall potential for harm and benefit to research subjects rather than an apparent inconsistency between any particular aspect of a research plan and a specific ethical guideline. For example, full disclosure of "what is really going on" in an experimental study is unnecessary if subjects are unlikely to be harmed. Nevertheless, researchers should make every effort to

foresee all possible risks and to weigh the possible benefits of the research against these risks. Researchers should consult with individuals with different perspectives to develop a realistic risk–benefit assessment and should try to maximize the benefits to, as well as minimize the risks for, subjects of the research (Sieber 1992:5–108).

Ultimately, these decisions about ethical procedures are not just up to you, as a researcher, to make. Your university's IRB sets the human subjects protection standards for your institution and may require that you submit any research proposal to the IRB for review. So I leave you with the instruction to review the human subjects guidelines of the ASA or other professional associations in your field, consult your university's procedures for the conduct of research with human subjects, make human subjects protections an important part of your research proposal, and then proceed according to the appropriate standards.

You should now also understand how important it is to plan all aspects of your research systematically and to present this plan in a formal research proposal. Human subjects protections are a vital part of any research plan that includes the collection of data from living people, but writing a research proposal is itself a way to protect yourself from the many problems that will arise if you do not think through carefully each stage of the research process before you start to collect data.

You are now ready to begin planning the particulars of a research project.

KEY TERMS

Belmont Report (p. 80)
Beneficence (p. 80)
Certificate of Confidentiality (p. 92)
Code of Ethics (American Sociological Association) (p. 81)
Conflict of interest (p. 85)
Debriefing (p. 89)
Deception (p. 90)
Federal Policy for the Protection of Human Subjects (Common Rule) (p. 81)
Health Insurance Portability and Accountability Act (HIPAA) (p. 92)
Institutional review board (IRB) (p. 98)
Justice (p. 80)
Milgram's obedience experiments (p. 79)
Nuremberg War Crime Trials (p. 78)
Office for Protection From Research Risks, National Institutes of Health (p. 98)
Process responsiveness (p. 100)
Respect for persons (p. 80)
Revised Common Rule
Tearoom Trade (p. 89)
Tuskegee Study of Untreated Syphilis in the Negro Male
Zimbardo prison simulation study (p. 87)

HIGHLIGHTS

LO 3.1 Discuss the human rights violations that led to the Belmont Report and subsequent regulations.

The Nuremberg War Crime Trials exposed the horrific medical experiments conducted by Nazi doctors and others in the name of "science."

The Tuskegee Study of Untreated Syphilis in the Negro Male was deceptive in that participants were told they were treated for "bad blood" (i.e., syphilis) whether they had syphilis or not; it harmed the participants because they were not offered effective treatment for syphilis after it became available, and participants were not asked for their consent to be studied.

Stanley Milgram's research on obedience to authority and Philip Zimbardo's prison simulation experiment both raised concerns about harm to subjects and deception.

Laud Humphrey's secretive observations and interviews under false pretenses of men who engaged in sex with other men in public bathrooms violated the principles of honesty and openness, informed consent, and avoiding deception.

LO 3.2 Judge the extent to which a research design satisfies each of the ethical principles for the protection of human subjects

Research should cause no harm to subjects.

Participation in research should be voluntary; therefore, subjects must give their informed consent to participate in the research and researchers must disclose their identity.

Researchers should avoid deception, except in limited circumstances.

Anonymity or confidentiality must be maintained for individual research participants unless it is waived voluntarily and explicitly.

Scientists must consider the uses of a research project so that its benefits outweigh any foreseeable risks.

LO 3.3 Compose an informed consent form.

The language of the consent form must be clear and understandable to the research participants and yet sufficiently long and detailed to explain what will actually happen in the research.

The new federal human subject regulations implemented fully in 2019 include a requirement for an initial concise explanation in consent forms that is intended to improve participant understanding.

Key information is to include identification of the consent as involving voluntary participation in research; the purposes, expected duration, and procedures of the research; and reasonably foreseeable risks and benefits; as well as available alternative treatments (if relevant).

LO 3.4 Defend the institutional review board requirement for informed consent in a research study.

An institutional review board is a group of organizational and community representatives required by federal law to review the ethical issues in all proposed research that is federally funded, involves human subjects, or has any potential for harm to human subjects.

The IRB has the authority to require changes in a research protocol or to refuse to approve a research protocol if it deems human subject protections inadequate.

IRB members review and approve informed consent forms to ensure that potential benefits of the research outweigh the risks.

LO 3.5 Create a detailed outline for each section of a research proposal.

An introductory statement of the research problem, in which you clarify what it is that you are interested in studying

A literature review, in which you explain how your problems and plans build on what has already been reported in the literature on this topic

A methodological plan, detailing just how you will respond to the mix of opportunities and constraints you face

An ethics statement, identifying human subject issues in the research and how you will respond to them in an ethical fashion

A statement of limitations, reviewing the potential weaknesses of the proposed research design and presenting plans for minimizing the consequences

DISCUSSION QUESTIONS

1. You probably know some people refused COVID-19 vaccines because they said they did not want to be used as "guinea pigs" by researchers. Some explicitly explained this attitude as reflecting the way that participants in the Tuskegee research were used. Yet thousands of people around the world participated in COVID-19 vaccine trials, and it is only because they did so that health care leaders were able to determine which vaccines were safe and effective. What strategies would you suggest for reducing fear of participating in research sponsored by organizations and researchers with appropriate expertise and oversight? Which ethical standards would you suggest emphasizing in recruiting participants?
2. Should social scientists be permitted to conduct replications of Milgram's obedience experiments? Zimbardo's prison simulation? Can you justify such research as permissible within the current ASA ethical standards? If not, do you believe that these standards should be altered to permit Milgram-type research?
3. How do you evaluate the current ASA ethical code? Is it too strict or too lenient or just about right? Are the enforcement provisions adequate? What provisions could be strengthened?
4. Why does unethical research occur? Is it inherent in science? Does it reflect "human nature"? What makes ethical research more or less likely?
5. Does debriefing solve the problem of subject deception? How much must researchers reveal after the experiment is over as well as before it begins?

SPSS EXERCISES

1. Consider three variables in the GSS2020 survey: helpblk_2, income16_2, and owngun_2. Review the corresponding variable labels in the "Variable View" in the GSS2020 file you are using so that you know what questions were used to measure these variables. Are there any ethical issues involved in asking these questions? Do you imagine that some respondents would be more likely to give untruthful answers to these questions? Explain your answers.
2. Examine the cross-tabulation between political views and support for the right to abortion after a rape and also support for allowing an antireligionist to speak (Analyze?Descriptive Statistics?Crosstabs? (Rows=abrape_2, spkath_2; Columns=polviews_2; Cells/Percentages=columns). Summarize what you have learned about the association between political conservatism and feelings about abortion and allowing an antireligionist to speak.

SECTION II

FUNDAMENTALS OF SOCIAL RESEARCH

Chapter 4 Conceptualization and Measurement 110

Chapter 5 Sampling and Generalizability 150

Chapter 6 Research Design and Causation 193

©iStockphoto.com/mauro_grigollo

4 CONCEPTUALIZATION AND MEASUREMENT

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

4.1 4.1Explain why the conceptualization of terms we use in our research is important

4.2 Discuss the relationship between conceptualization and operationalization

4.3 Illustrate key forms of measurement, including surveys, observations, unobtrusive measures, and published data

4.4 Produce examples of the four levels of measurement for a given concept

4.5 Compare the different approaches to measurement validity and reliability and recommend ways to improve validity and reliability

Tolerance of others has been in short supply in the last decade in many countries, including the United States. Whether in protests about immigration, demonstrations supporting or opposing reduced funding for the police, or campaigns for or against mask requirements and vaccination mandates, intense feelings, loud voices, and charges of prejudice toward other groups can make it seem that democratic give-and-take is being replaced by extreme intolerance for other people and other beliefs. It is no surprise that the United Nations and other international organizations that intentionally bring together people from many different backgrounds emphasize tolerance as a foundation for their organizational success and for peace between nations (United Nations 1945).

Whether your goal is to learn whether the level of tolerance in society is declining, to develop social policies that improve tolerance, or to simply identify places where you'd rather live, at some point, you are likely to think more about the meaning of tolerance and perhaps even read some of the related research literature. Perhaps you will even attempt to design your own study of tolerance. Every time you begin to review or design relevant research, you will have to answer two questions: (1) What is meant by *tolerance* in this research (conceptualization)? and (2) How was tolerance measured (operationalization)? Both types of questions must be answered when we evaluate prior research, and both types of questions must be kept in the forefront when we design new research. Only when we conclude that a study used valid measures of its key concepts can we have some hope that its conclusions are valid.

In this chapter, I first address the issue of conceptualization, using tolerance and other concepts as examples. I then focus on measurement, reviewing first how indicators of tolerance and several other concepts have been constructed using such operations as questions, observations, and less direct and obtrusive measures. Next, I discuss the different possible levels of measurement and methods for assessing the validity and reliability of measures. The final topic to consider is the unique insights that qualitative methods can add to the measurement process. By the chapter's end, you should have a good understanding of conceptualization and measurement, the first of the three legs on which a research project's validity rests.

CONCEPTS

Although it may seem like a good thing, what does tolerance really mean? That we put up with people we dislike? That we have a "live and let live" attitude toward others who think or behave in ways that are very different from our own? That we may admire others who we perceive as different in some way? Before we can investigate how tolerant people are or whether there has been a decline in the level of tolerance in any group or society, we must specify what we mean by the term. In their review of literature that suggests the need for a "new approach to the study of tolerance," Mikael Hjerm, Maureen A. Eger, Andrea Bohman, and Filip Fors Connolly (2020) at Umeå University in Sweden show that many researchers have assumed "tolerance" of others is simply the opposite of feelings of "prejudice" toward others. In contrast, Hjerm et al. (2020) argue that tolerance and prejudice should be understood as very different concepts. They begin their research on tolerance in Sweden and other nations by developing a different concept of tolerance and then specifying several ways to measure that concept.

Mikael Hjerm and his colleagues propose that tolerance should be defined as an orientation to diversity, not in relation to specific disliked groups. Specifically, Hjerm et al. (2020:903) define tolerance as "a value orientation towards difference." The more tolerant people are, in terms of this definition, the more they accept difference, respect difference, and appreciate difference. In this way, tolerance is defined as different than simply putting up with disliked groups (or not being prejudiced toward them).

Which definition of *tolerance* makes more sense to you? We can't say that only one definition of *tolerance* is "correct" or even that one is "better." What we can say is that we need to specify what we mean when we use the term. We also must be sure others know what definition we are using. And of course, the definition should be useful for our purposes: A definition based solely on feelings of prejudice will not be useful if we think a community can thrive even with some prejudiced residents if they are tolerant of others.

We call tolerance a **concept**—a mental image that summarizes a set of similar observations, feelings, or ideas. To make that concept useful in research (and even in ordinary discourse), we must define it. Many concepts are used in everyday discourse without consistent definition, sometimes definitions of concepts are themselves the object of intense debate, and the meanings of concepts may change over time. For example, when we read a *New York Times* article (Stille 2000) announcing a rise in the "social health" of the United States, after a precipitous decline in the 1970s and 1980s, we don't know whether we should feel relieved or disinterested. In fact, the authorities on the subject didn't even agree about what the term *social health* meant: lessening of social and economic inequalities (Marc Miringoff) or clear moral values (William J. Bennett). Most agreed social health has to do with "things that are not measured in the gross national product" and that it is "a more subtle and more meaningful way of measuring what's important to [people]" (Stille 2000, A19), but the sparks flew over whose **conceptualization** of social health would prevail.

Isabel Wilkerson (2020) proposes in *Caste: The Origins of Our Discontents* that the concept of caste can be used to better describe and understand the role of race in the U.S. stratification system. Although more commonly used to refer to India's traditional explicit caste system, Wilkerson (2020:17) defines "caste" in a manner to be applied to systems that do not use the term:

> A caste system uses rigid, often arbitrary boundaries to keep the ranked groupings apart, distinct from one another and in their assigned places. . . . A caste system endures because

> it is often justified as divine will, originating from sacred text or the presumed laws of nature, reinforced throughout the culture and passed down through the generations.

She goes on to distinguish concepts of caste and race as "neither synonymous or mutually exclusive" and terms "race, in the United States [as] the visible agent of the unseen force of caste" (Wilkerson 2020:19).

Concepts such as tolerance, social health, and caste require an explicit definition before they are used in research because we cannot be certain all readers will share a particular definition or that the current meaning of the concept is the same as it was when previous research was published. It is especially important to clearly define any concepts that are abstract or unfamiliar. When we refer to concepts such as *social control, social health, caste*, or *anomie,* we cannot count on others knowing exactly what we mean. Even experts may disagree about the meaning of frequently used concepts if they have based their conceptualizations on different theories. That's okay. The point is not that there can only be one definition of a concept but that we must specify clearly what we mean when we use a concept, and we must expect others to do the same.

Conceptualization in Practice

If we are to do an adequate job of conceptualizing, we must do more than just think up some definition, any definition, for our concepts (Goertz 2006). We need to identify what we think is important about the phenomenon that interests us. We should also turn to social theory and prior research to review appropriate definitions. We should understand how the definition we choose fits within the theoretical framework guiding the research and what assumptions underlie this framework. We may decide the concept has several dimensions, or subconcepts, that should be distinguished.

Tolerance

What ideas and behaviors should we associate with the concept *tolerance?* Is someone displaying tolerance if they believe in the right of free speech for others in groups or social categories they don't like? Does being tolerant mean a person feels comfortable interacting with people even when they have prejudicial feelings toward them? That they are friendly with all sorts of people? Or simply that they don't harbor prejudice toward others? Hjerm and his colleagues (2020) review these and other ways of conceptualizing tolerance and decide that "acceptance of difference" should be considered the defining aspect of tolerance. In this way, they distinguish the concept of tolerance from the concept of prejudice. They argue that someone who does not dislike particular other groups is not prejudiced but may still be intolerant if they don't accept, respect, and appreciate the existence of diversity between people.

In this way, conceptualizing one concept often requires defining it in relation to other concepts. Conceptualization of one concept may also require defining other terms used to define that concept. If our audience does not already have a shared understanding of the meaning of terms like "acceptance," "respect," and "appreciation"—terms used by Hjerm et al. (2020) to identify different dimensions of the concept of diversity, we must also define these related terms before we are finished with the process of defining the concept of tolerance.

Poverty

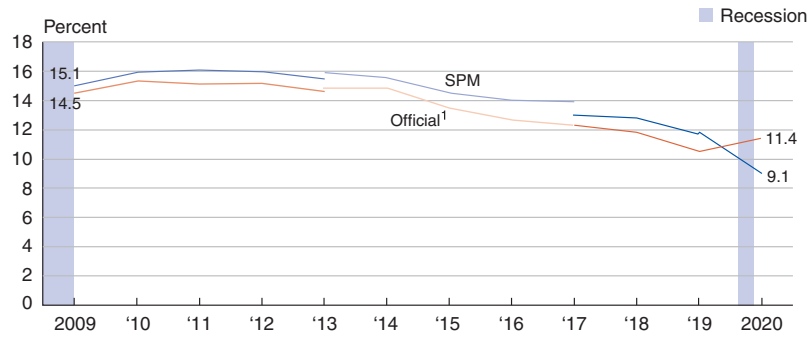
Decisions about how to define a concept reflect the theoretical framework that guides the researchers. For example, the concept *poverty* has always been somewhat controversial because

different conceptualizations of poverty lead to different estimates of its prevalence and different social policies for responding to it.

Most of the statistics about the U.S. poverty rate that you see in the news reflect a conception of poverty that was formalized by Mollie Orshansky of the Social Security Administration in 1965 and subsequently adopted by the federal government and many researchers (Putnam 1977). She defined *poverty* in terms of what is called an *absolute* standard, based on the amount of money required to purchase an emergency diet that is estimated to be nutritionally adequate for about two months. The idea is that people are truly poor if they can just barely purchase the food they need and other essential goods. This poverty standard is adjusted for household size and composition (number of children and adults), and the minimal amount of money needed for food is multiplied by three because a 1955 survey indicated that poor families spend about one third of their incomes on food (Orshansky 1977).

Does this sound straightforward? As is often the case with important concepts, the meaning of an absolute poverty standard has been the focus of a vigorous debate (Eckholm 2006, A8). Although the traditional definition of absolute poverty only accounts for a family's cash income, some observers argue noncash benefits that low-income people can receive, such as food stamps, housing subsidies, and tax rebates, should be added to cash income before the level of poverty is calculated. Douglas Besharov of the American Enterprise Institute terms this approach "a much needed corrective" (Eckholm 2006:A8). The impact of considering government benefits in the calculation of the poverty rate became starkly apparent during the coronavirus pandemic, when Congress spent billions of dollars to expand benefits and send direct checks to most Americans (Casselman and Smialek 2021; Mandel and Ludwick 2021). As a result, although the official poverty rate increased as millions lost their jobs, the heightened benefits succeeded in lowering the poverty rate calculated after taking these benefits into account (Figure 4.1).

FIGURE 4.1 ■ Official Poverty Rate and Poverty Rate Including Benefits



[1]Includes unrelated individuals under the age of 15.
Notes: The Supplemental Poverty Measure (SPM) estimates for 2019 and 2020 reflect the implementation of revised SPM methodology. More information is available in the report appendix. The data for 2017 and beyond reflect the implementation of an updated processing system. The data for 2013 and beyond reflect the implementation of the redesigned income questions. More information on confidentially protection. sampling error, nonsampling error, and definitions is available at <https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar21.pdf>.

Source: Fox, L. E., and Burns, K. 2021. The Supplemental Poverty Measure: 2020. U.S. Department of Commerce, U. S. Census Bureau (September 2021; P60–275), census.gov.

But some social scientists have proposed increasing the absolute standard for poverty so it reflects what a low-income family must spend to maintain a "socially acceptable standard of living" that allows for a telephone, house repairs, and decent clothes (Uchitelle 1999). A new "Multidimensional Poverty Index" (MPI) to aid international comparisons considers absolute deprivations in health, education, and living standards (Alkire et al. 2011). Others argue the persistence of poverty should be considered, so someone who is poor for no more than a year, for example, is distinguished from someone who is poor for many years (Walker, Tomlinson, and Williams 2010:367–368). Any change in the definition of poverty will change eligibility for government benefits, such as food stamps and Medicaid, so the feelings about this concept run deep.

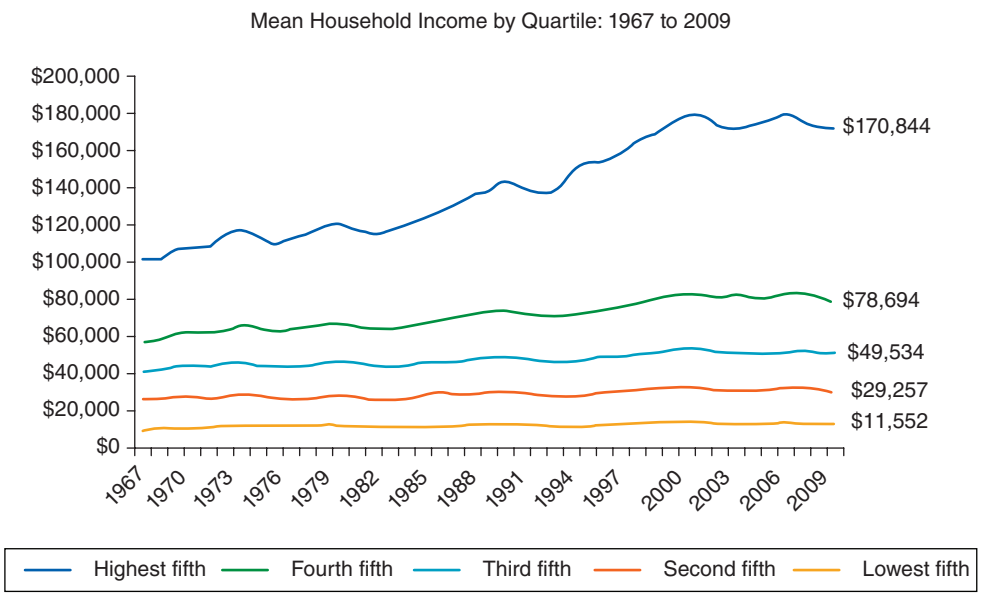
Some social scientists disagree altogether with the absolute standard and have instead urged adoption of a *relative* poverty standard (see Figure 4.2). They identify the poor as those in the

FIGURE 4.2 ■ Absolute, Relative, and Subjective Poverty Standards

Absolute Standard

Subjective Standard

Relative Standard



Based on Giovanni Vecchi, Universita di Roma "Tor Vergata," Poverty Lines. Bosnia and Herzegovina Poverty Analysis Workshop, September 17–21, 2007.

lowest fifth or tenth of the income distribution or as those having some fraction of the average income. The idea behind this relative conception is that poverty should be defined in terms of what is normal in a given society at a particular time. "For example, while a car may be a luxury in some poor countries, in a country where most families own cars and public transportation is inadequate, a car is a basic necessity for finding and commuting to work" (Mayrl et al. 2004:10).

Some social scientists prefer yet another conception of poverty. With the *subjective* approach, *poverty* is defined as what people think would be the minimal income they need to make ends meet. Of course, many have argued this approach is influenced too much by the different standards people use to estimate what they "need" (Ruggles 1990:20–23).

Which do you think is a more reasonable approach to defining poverty: some type of absolute standard, a relative standard, or a subjective standard? Be careful here: Conceptualization has consequences! Research using the standard absolute concept of poverty indicated the percentage of Americans in poverty declined by 1.7% in the 1990s, but use of a relative concept of poverty led to the conclusion that poverty increased by 2.7% (Mayrl et al. 2004:10). No matter which conceptualization we decide to adopt, our understanding of the concept of poverty will be sharpened after we consider these alternative definitions.

RESEARCH IN THE NEWS
HAVE AMERICANS WARMED TO CALLS TO "DEFUND THE POLICE"?

"Defund the police" became a household phrase in the wake of George Floyd's killing by Minneapolis police officers in May 2020. During the same period, a Kaiser Family Foundation poll found that about three quarters of Americans said police violence against the public was a problem and supported increased transparency about police misconduct, while some activists sought to translate this sentiment into legislative action to reduce funding for police. However, an Associated Press/NORC poll found only 25% of Americans supported reducing funding for police departments. What did respondents think "reducing funding for police departments" means? When a Fox News poll asked a more specific question about shifting funds from police departments to "mental health, housing, and other social services," 41% of voters agreed. When a Siena College poll asked New York City residents whether mental health professionals should come along with police officers responding to calls related to homelessness, drug addiction, or mental illness, 90% agreed. A PerryUdem survey found many people interpreted "defund the police" as a call to move funds from the police to other community services, some thought it meant just taking money away from the police, and many were not sure what it meant. So the wording of the specific questions used to measure support for defunding the police made a big difference in the level of support identified.

For Further Thought?

1. What question would you recommend to measure support for the concept of "defunding the police"? How would you define the concept of "defunding the police"?
2. Can we depend on self-report measures of attitudes toward funding the police? What measurement challenges can you think of?

News source: Russonello, Giovanni. 2020. "Have Americans Warmed to Calls to 'Defund the Police'?" *The New York Times*, July 3 (updated online August 4).

FROM CONCEPTS TO INDICATORS

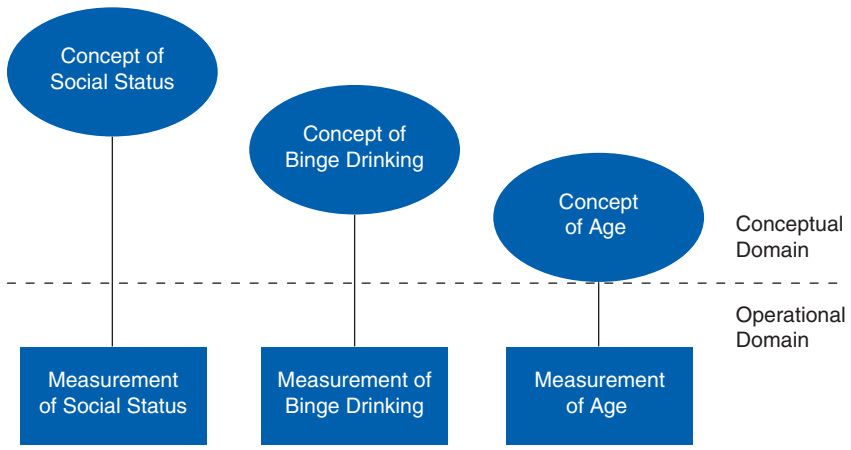
Identifying the concepts we will study, specifying dimensions of these concepts, and defining their meaning only begin the process of connecting our ideas to concrete observations. If we are to conduct empirical research involving a concept, we must be able to distinguish it in the world around us and determine how it may change over time or differ between persons or locations. **Operationalization** involves connecting concepts to measurement operations. You can think of it as the empirical counterpart of the process of conceptualization. When we *conceptualize*, we specify what we mean by a term. When we *operationalize*, we identify specific measurements we will use as **indicators** of that concept in empirical reality. Researchers also find the process of figuring out how to measure a concept helps improve their understanding of what the concept means (Bartholomew 2010:457). Improving conceptualization and improving operationalization go hand in hand.

Concepts vary in their level of abstraction, and this, in turn, affects how readily we can specify the indicators pertaining to the concept. We may not think twice before we move from a conceptual definition of *age* as time elapsed since birth to the concrete indicator "years since birth." Binge drinking is also a relatively concrete concept, but it requires a bit more thought (see Figure 4.3). Most researchers define *binge drinking* conceptually as heavy episodic drinking and operationally as drinking five or more drinks in a row (for men) (Wechsler et al. 2002:205). That's pretty straightforward, although we still need to specify the questions that will be used to determine the frequency of drinking.

An abstract concept such as *social status* may have a clear role in social theory but still have different meanings in different social settings. Indicators that pertain to social status may include level of esteem in a group, extent of influence over others, level of income and education, or number of friends. It is very important to specify what we mean by an abstract concept such as social status in a particular study and to choose appropriate indicators to represent this meaning.

You have already learned in Chapter 2 that variables are phenomena that vary. Usually, the term *variable* is used to refer to some specific aspect of a concept that varies and for which we then have to select even more concrete indicators. For example, research on the *concept* of social

FIGURE 4.3 ■ Varying Distances Between Concepts and Measures



Source: Adapted from Viswanathan, Madhu. 2005. *Measurement Error and Research Design*. Thousand Oaks, CA: Sage (p. 7).

support might focus on the *variable* "level of perceived support," and we might then select as our *indicator* the responses to a series of statements about social support, such as this one from S. Cohen et al.'s (1985) social support index, the "Interpersonal Support Evaluation List": "If I needed a quick emergency loan of $100, there is someone I could get it from" (p. 93). Identifying the variables we will measure moves us closer to developing our specific measurement procedures. I give more examples in the next section.

The term *variable* is sometimes used interchangeably with the term *indicator*, however. This means you might find "crime rate" or "importance of extrinsic rewards" being termed as either variables or indicators. Sometimes the term *variable* is used to refer to phenomena that are more abstract, such as "alienation" or "social capital." You might hear one researcher referring to *social support* as one of the important concepts in a study, another referring to it as a variable that was measured, and another calling it an indicator of group cohesion. The important thing to remember is that we need to define clearly the concepts we use and then develop specific procedures for indicating variation in the variables related to these concepts.

Bear in mind that concepts don't necessarily vary. For example, gender may be an important concept in a study of influences on binge drinking, but it isn't a variable in a study of members of an all-male fraternity. When we explain excessive drinking in the fraternity, we might attach great importance to the all-male fraternity subculture. However, because gender doesn't vary in this setting, we won't be able to study differences in binge drinking between male and female students. So gender will be a **constant**, not a variable, in this study (unless we expand our sample to include members of both sororities and fraternities or perhaps the general student population).

Operationalizing the Concept of Race

Race is an important concept in social research. In research applications as in everyday life, the concept of race is often treated as if it is an obvious distinction of several categories based on physical appearance (in turn related to ancestry). But in fact, what people mean by the concept of *race* has varied over time and differs between countries. These inconsistencies become clear only when we examine how the concept of race has been operationalized in specific questions.

Repeated changes in questions about race in the decennial U.S. Census reflect social and political pressures (Snipp 2003:565–567). Race was not assessed directly in the censuses of 1790, 1800, or 1810, but enslaved Africans and Native Americans were distinguished from others. In 1820, 1830, and 1840, *color* was distinguished; then, as racial consciousness heightened before the Civil War, the 1850 census added the category of *Mulatto* to identify people of mixed-race parentage. In response to concerns with increasing immigration, *Chinese* and *Indian* (Asian) persons were distinguished in 1860 and then *Japanese* was added to the list in 1870. In the 1890 census, *Octoroons* and *Quadroons* were distinguished as different categories of Mulatto. The infamous 1896 U.S. Supreme Court decision *Plessy v. Ferguson* reflected the victory of "Jim Crow" legislation in the South and defined as Black any person who had as much as one Black ancestor. (Homer Plessy was a Louisiana shoemaker whose skin was white, but he was told he could not ride in the "whites only" section of the train because he had one Black ancestor.) By 1920, the U.S. Census reflected this absolute distinction between persons judged Black and white by dropping the distinctions involving mixed-race ancestry. In 1930 and 1940, *Mexican* was distinguished, but in 1950, political pressure led to this category's being dropped as an ethnic minority; instead, Mexicans were treated as *white* (Snipp 2003:568–569).

By the late 1950s, the civil rights movement began to influence the concept of race as used by the U.S. Census. In 1960, the census shifted from assessing race on the basis of physical appearance to self-identification (Snipp 2003:569–570). As one result, the number of people identified

as Native Americans rose dramatically. More important, this shift reflected a change in thinking about the concept of race as reflecting primarily physical appearance and instead indicating cultural identification. In the 1970s, the Federal Interagency Committee on Education (FICE) established an ad hoc committee that led the 1980 U.S. Census (and all federal agencies) to use a five-category distinction: (1) American Indians and Alaska Natives, (2) Asians and Pacific Islanders, (3) non-Hispanic Blacks, (4) non-Hispanic whites, and (5) Hispanics (Snipp 2003:572–574). In that census, Spanish/Hispanic origin or descent was asked as a question distinct from the question about race (U.S. Census Bureau 1981:3).

But the new concept of race reflected in these five categories only led to new complaints. Some parents in "mixed-race" marriages insisted they should be able to classify their children as multiracial (Snipp 2003:575–576)—in opposition to some civil rights leaders concerned with diluting the numbers of Americans designated as Black (Holmes 2001b:WK1). The chair of the Census Bureau's Hispanic advisory committee complained, "We don't fit into the categories that the Anglos want us to fit in" (Swarns 2004:A18). Sociologist Orlando Patterson (1997) and many others argued that the concept of ethnicity was more meaningful than race. As a result of these and other complaints, a new federal task force developed a new approach that allowed respondents to designate themselves as being of more than one race. The resulting question about race reflected these changes as well as increasing distinctions within what had been the Asian and Pacific Islanders category. In the 2010 census, 3% of Americans identified themselves as multiracial (Humes, Jones, and Ramirez 2011). An official census report after the 2010 Census also included this caveat about the definition:

> The race categories included in the census questionnaire generally reflect a social definition of race recognized in this country and are not an attempt to define race biologically, anthropologically, or genetically. In addition, it is recognized that the categories of the race question include race and national origin or sociocultural groups. (Humes, Jones, and Ramirez, 2011:2).

In preparation for the 2020 Census, the Census Bureau tested different race and ethnicity questions with a sample of 1.2 million housing units in the United States, including Puerto Rico. (Mathews, 2017). The decisions made after this "National Content Test" led to questions in the 2020 U.S. Census that asked respondents to check one or more "race" categories and, in most cases, to add details about their specific national origins (such as Jamaican, Nigerian, Cambodian) (see Figure 4.4). Almost one in five (18.7%) classified themselves as Hispanic or Latino (regardless of race), 12.4% as Black or African American, 6.0% as Asian, 1.1% as Native American or Alaska Native, 61.6% as white (not Hispanic or Latino), 10.2% as "two or more races," and 8.4% as "some other race" (U.S. Census 2021 https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html).

The concept of race also varies internationally, so any research involving persons in or from other countries may need to use a different definition of race. For example, Darryl Fears, former director of the Brazilian American Cultural Institute (in Washington, D.C.), explains how social conventions differ in Brazil: "In this country [the U.S.], if you are not quite white, then you are black." But in Brazil, "If you are not quite black, then you are white" (Fears 2002:A3). In Mexico, the primary ethnic distinction is between indigenous and nonindigenous residents, without a clear system of categorization based on skin color. Nonetheless, survey research indicates a marked preference for whiter skin and "profound social stratification by skin color" (Villarreal 2010:671). The conception and operationalization of race, then, varies with place.

FIGURE 4.4 ■ The U.S. Census Bureau Ethnicity and Race Questions in 2020

8. Is Person 1 of Hispanic, Latino, or Spanish origin?
   - ☐ No, not of Hispanic, Latino, or Spanish origin
   - ☐ Yes, Mexican, Mexican Am., Chicano
   - ☐ Yes, Puerto Rican
   - ☐ Yes, Cuban
   - ☐ Yes, another Hispanic, Latino, or Spanish origin - *Print, for example, Salvadoran, Dominican, Colombian Guatamalan, Spaniard, Ecuadorian, etc.* →

9. What is Person 1's race?
   Mark ☒ *one or more boxes* ***AND*** *print origins.*
   - ☐ White - *Print, for example, German, Irish, English, Italian Lebanese. Egyptian. etc.* →
   - ☐ Black or African Am. - *Print for example. African American, Jamaican, Haitian Nigerian, Ethiopian, Somali, etc.* →
   - ☐ American Indian or Alaska Native - *Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Govermment Nome Eskimo Community etc.* →
   - ☐ Chinese
   - ☐ Filipino
   - ☐ Asian Indian
   - ☐ Vietnamese
   - ☐ Korean
   - ☐ Japanese
   - ☐ Other Asian *Print, for example, Pakistani Cambodian, Hmong etc.* →
   - ☐ Native Hawaiian
   - ☐ Samoan
   - ☐ Chamorro
   - ☐ Other Pacific lalander *Print, for example, Tongan, Fijian, Marshallese, etc.* →
   - ☐ Some other race - *Print race or orgin* →

Source: U.S. Census Bureau, 2020 Census Questionnaire.

Operationalizing Perceived Threat From Immigrants

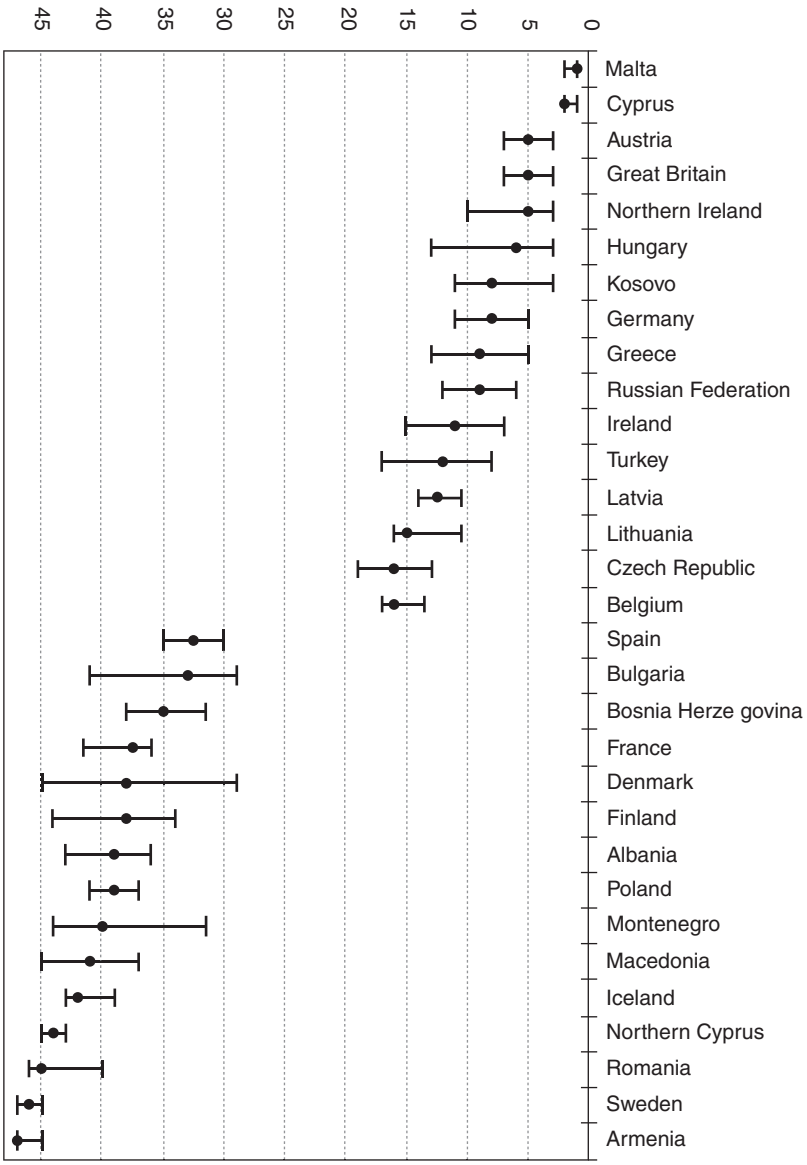
Conflict over whether to welcome or expel migrants has riven many countries in Europe and in the United States, as exemplified in the 2016 and 2020 presidential elections in the United States. The concept of perceived threat from immigrants has therefore become central to many research questions, and many researchers have developed methods for operationalizing this concept. For example, Marco Marozzi (2016) investigated perceived level of socioeconomic threat from immigrants in 47 European countries and regions. He used responses to statements about immigrants in the European Values Study between March 2008 and March 2010 to construct his operational measure of perceived socioeconomic threat from immigrants.

Each of the statements was offered as part of a pair to respondents in the European Values Study survey. One member of a pair would be a statement like "Immigrants do not take jobs away from natives in a country," and the other would then be a statement expressing the opposite sentiment, like "Immigrants take jobs away from natives in a country." Respondents were asked to indicate on a scale from 1 to 10 whether they tended to agree more with the first statement or

the second one. Other paired statements rated in this way included perceived effects of immigrants on cultural life, on crime, on the welfare system, and on society as a whole. The ratings for each of five pairs were then averaged to indicate how much the respondent endorsed the statements that favored immigrants. This average rating was the empirical indicator used to operationalize the concept of perceived threat from immigrants.

Figure 4.5 shows the countries with the four highest and lowest average scores based on these ratings. Respondents in Armenia, Sweden, Romania, and Northern Cyprus perceived

FIGURE 4.5 ■ Perceived Threat From Immigrants in Selected European Countries



Source: Based on Marozzi, Marco. 2016. "Construction, Robustness Assessment and Application of an Index of Perceived Level of Socio-Economic Threat from Immigrants: A Study of 47 European Countries and Regions." *Social Indicators Research* 128:430.

immigrants as least threatening, while those in Great Britain, Austria, Cyprus, and Malta rated immigrants as most threatening. Marozzi (2016) then considers possible explanations for the inter-country differences. He notes that Malta, the country where the perceived threat indicator score is the highest, is a small island in the middle of the Mediterranean where many immigrants arrive from northern Africa, while Sweden, which is second to last on the list, "is a well-known tolerant country" (Marozzi 2016:431). But the ranking of some other countries raises new questions, such as why Cyprus and North Cyprus are at opposite ends of the ranking. In this way, operationalizing the concept of perceived threat from immigrants provides a foundation for explanatory analyses and raises new questions for additional research.

The Inductive Approach to Conceptualization

Qualitative research projects usually take an inductive approach to the process of conceptualization. In an inductive approach, concepts emerge from the process of thinking about what has been observed, compared with the deductive approach that I have just described, in which we develop concepts on the basis of theory and then decide what should be observed to indicate that concept. So instead of deciding in advance which concepts are important for a study, what these concepts mean, and how they should be measured, if you take an inductive approach, you will begin by recording verbatim what you hear in intensive interviews or see during observational sessions. You will then review this material to identify important concepts and their meaning for participants. As your understanding of the participants and social processes develops, you may refine your concepts and modify your indicators.

You will learn more about qualitative research in Chapter 10, but an example here will help you understand the qualitative measurement approach. For several months, Darin Weinberg (2000) observed participants in three drug abuse treatment programs in southern California. He was puzzled by the drug abuse treatment program participants' apparently contradictory beliefs—that drug abuse is a medical disease marked by "loss of control" but also that participation in a therapeutic community can be an effective treatment. He discovered that treatment participants shared an "ecology of addiction" in which they conceived of being "in" the program as a protected environment, whereas being in the community was considered being "out there" in a place where drug use was inevitable—in "a space one's addiction compelled one to inhabit" (Weinberg 2000:609).

> I'm doin' real, real bad right now. . . . I'm havin' trouble right now staying clean for more than two days. . . . I hate myself for goin' out and I don't know if there's anything that can save me anymore. . . . I think I'm gonna die out there. (Weinberg 2000:609)

Participants contrasted their conscientiousness while in the program with the personal dissolution of those out in "the life."

So Weinberg developed the concepts of *in* and *out* inductively, in the course of the research, and he identified indicators of these concepts at the same time in the observational text. He continued to refine and evaluate the concepts throughout the research. Conceptualization, operationalization, and validation were ongoing and interrelated processes.

MEASUREMENT

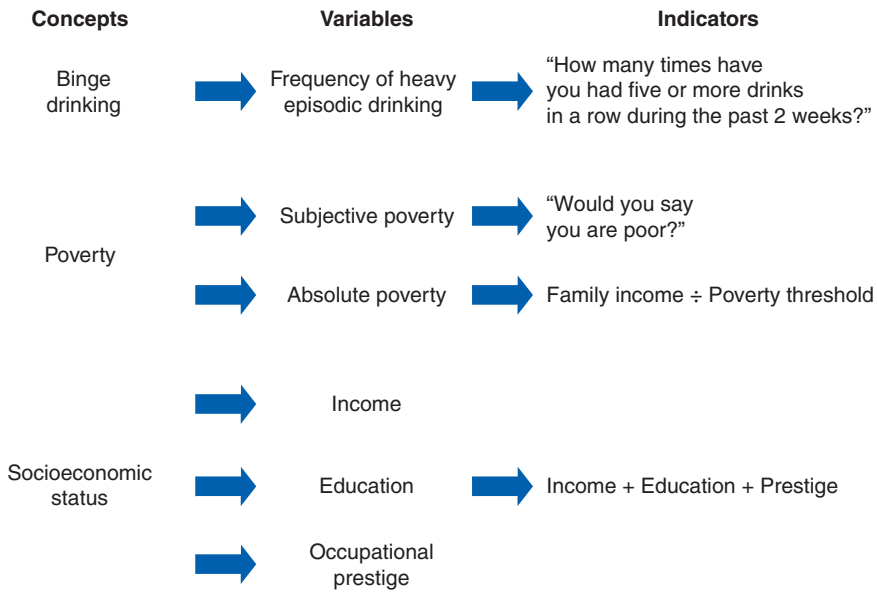
The deductive researcher proceeds from defining concepts in the abstract (conceptualizing) to identifying variables to measure, and finally to developing specific measurement procedures. **Measurement** is the "process of linking abstract concepts to empirical indicants" (Carmines and

Zeller 1979:10). The goal is to achieve measurement validity, so the measures, or indicators, must measure the variables they are intended to measure.

Figure 4.6 represents the operationalization process in three studies. The first researcher defines her concept, *binge drinking*, and chooses one variable—frequency of heavy episodic drinking—to represent it. In the Monitoring the Future survey (Jang, Patrick, and Schuler 2017), the single question, or indicator, used was "How many times have you had five or more drinks in a row during the past 2 weeks?" Binge drinking was defined as a response greater than 5. A researcher may define the concept *poverty* as having two aspects or dimensions: subjective poverty and absolute poverty. Subjective poverty is measured with responses to the survey question "Would you say you are poor?" Absolute poverty is measured by comparing family income to the poverty threshold. A researcher may operationalize the concept *socioeconomic status* by a position on a combination of responses to three measured variables: income, education, and occupational prestige.

Social researchers have many options for operationalizing concepts. Measures can be based on activities as diverse as asking people questions, reading judicial opinions, observing social interactions, coding words in books, checking census data tapes, enumerating the contents of trash receptacles, or drawing urine and blood samples. Experimental researchers may operationalize a concept by manipulating its value. For example, to operationalize the concept of *positive racial imagery*, in their study of its effect in electoral campaigns, Tasha S. Philpot and Kenneth M. Miller (2020) showed respondents a political ad that included one of two images—with one displaying inclusive racial imagery (see Figure 4.7). I will focus here on the operations of asking questions, observing behavior, using unobtrusive means of measuring people's behavior and attitudes, and using published data.

FIGURE 4.6 ■ Concepts, Variables, and Indicators



Constructing Questions

Asking people questions is the most common and the most versatile operation for measuring social variables. Most concepts about individuals can be defined in such a way that measurement

FIGURE 4.7A ■ Response to Imagery in Ads (a) Race-Neutral Image

©iStockphoto.com/DGLimages

FIGURE 4.7B ■ Response to Imagery in Ads (b) Racially Inclusive Image

©iStockphoto.com/SDI Productions

with one or more questions becomes an option. We associate questions with survey research, but questions are also often the basis of measures used in social experiments and in qualitative research. In this section, I introduce some options for writing single questions; in Chapter 8, I explain why single questions can be inadequate as measures and then examine measurement approaches using multiple questions to measure a concept.

Of course, even though questions are, in principle, a straightforward and efficient means to measure individual characteristics, facts about events, level of knowledge, and opinions of any sort, they can easily result in misleading or inappropriate answers. Memories and perceptions of

the events about which we might like to ask can be limited, and some respondents may intentionally give misleading answers. For these reasons, all questions proposed for a study must be screened carefully for their adherence to basic guidelines and then tested and revised until the researcher feels some confidence that they will be clear to the intended respondents and likely to measure the intended concept (Fowler 1995). Alternative measurement approaches will be needed when such confidence cannot be achieved.

Specific guidelines for reviewing survey questions are presented in Chapter 8; here, my focus is on the different types of questions used in social research.

Measuring variables with single questions is very popular. Public opinion polls based on answers to single questions are reported frequently in news articles and TV newscasts: "Do you favor or oppose U.S. policy . . . ?" "If you had to vote today, for which candidate would you vote?" Social science surveys also rely on single questions to measure many variables, for instance, "Overall, how satisfied are you with your job?" or "How would you rate your current health?"

Single questions can be designed with or without explicit response choices. The question that follows is a **closed-ended (fixed-choice) question** because respondents are offered explicit responses from which to choose. It has been used in Gallup polls and was included In a YouGov.org online survey of gun control that Alexandra Filindra and Noah Kaplan (2017) used to test differences in gun policy preferences among people who are Black, Latinx, and white.

> In general, do you feel that the laws covering the sale of firearms should be made more strict, less strict, or kept as they are now? (Mark one)
>
> ____ More strict
>
> ____ Less strict
>
> ____ Kept as they are now

Most surveys of large numbers of people contain primarily fixed-choice questions, which are easy to process with computers and analyze with statistics. With fixed-choice questions, respondents are also more likely to answer the questions the researcher really wants them to answer. Including response choices reduces ambiguity and makes it easier for respondents to answer. However, fixed-response choices can obscure what people really think if the choices do not match the range of possible responses to the question; many studies show that some respondents will choose response choices that do not apply to them simply to give some sort of answer (Peterson 2000:39).

Most important, response choices should be **mutually exclusive** and **exhaustive**, so that every respondent can find one and only one choice that applies to them (unless the question is of the "Check all that apply" format). To make response choices exhaustive, researchers may need to offer at least one option with room for ambiguity. For example, a questionnaire asking college students to indicate their school status should not use freshman, sophomore, junior, senior, and graduate student as the only response choices. Most campuses also have students in a "special" category, so you might add "Other (please specify)" to the five fixed responses to this question. If respondents do not find a response option that corresponds to their answer to the question, they may skip the question entirely or choose a response option that does not indicate what they are really thinking.

Open-ended questions, questions without explicit response choices, to which respondents write in their answers, are preferable when the range of responses cannot adequately be anticipated—namely, questions that have not previously been used in surveys and questions that are

asked of new groups—or when there are simply a long list of possible responses. Open-ended questions can also lessen confusion about the meaning of responses involving complex concepts. The 2020 U.S. Census included an open-ended national origins question in which respondents could print their country of origin.

In qualitative research, open-ended questions are often used to explore the meaning respondents give to abstract concepts. *Mental illness*, for example, is a complex concept that tends to have different meanings for different people. In a survey I conducted in homeless shelters, I asked the staff members whether they believed people at the shelter had become homeless because of mental illness (Schutt 1992). When given fixed-response choices, 47% chose "Agree" or "Strongly agree." However, when these same staff members were interviewed in depth, with open-ended questions, it became clear that the meaning of these responses varied among staff members. Some believed mental illness caused homelessness by making people vulnerable in the face of bad luck and insufficient resources:

> Mental illness [is the cause]. Just watching them, my heart goes out to them. Whatever the circumstances were that were in their lives that led them to the streets and being homeless I see it as very sad. . . . Maybe the resources weren't there for them, or maybe they didn't have the capabilities to know when the resources were there. It is misfortune. (Schutt 1992:7)

Other staff believed mental illness caused people to reject housing opportunities:

> I believe because of their mental illness that's why they are homeless. So for them to say I would rather live on the street than live in a house and have to pay rent, I mean that to me indicates that they are mentally ill. (Schutt 1992:7)

The decision to use closed-ended or open-ended questions can have important consequences for the information reported. Leaving an attitude or behavior off a fixed set of response choices is likely to mean that it is not reported. However, respondents to self-administered questionnaires are more likely to skip open-ended questions entirely (Schwarz 2010:48).

Making Observations

Observations can be used to measure characteristics of individuals, events, and places. The observations may be the primary form of measurement in a study, or they may supplement measures obtained through questioning.

Direct observations can be used as indicators of some concepts. For example, Albert Reiss (1971a) studied police interaction with the public by riding in police squad cars, observing police–citizen interactions, and recording their characteristics on a form. Notations on the form indicated variables, such as how many police–citizen contacts occurred, who initiated the contacts, how compliant citizens were with police directives, and whether police expressed hostility toward the citizens.

Using a different approach, psychologists Dore Butler and Florence Geis (1990) studied unconscious biases and stereotypes they thought might hinder the advancement of women and members of underrepresented communities in work organizations. In one experiment, discussion groups of male and female students were observed from behind one-way mirrors as group leaders presented identical talks to each group. The trained observers (who were not told what the study was about) rated the number of frowns, furrowed brows, smiles, and nods of approval as the group leaders spoke. (The leaders themselves did not know what the study was about.) Group participants made disapproving expressions, such as frowns, more often when the group

leader was a woman than when the leader was a man. To make matters worse, the more the women talked, the less attention they were given. Butler and Geis concluded there was indeed a basis for discrimination in these unconscious biases.

Direct observation is often the method of choice for measuring behavior in natural settings. Direct observation avoids the problems of poor recall and self-serving distortions that can occur with answers to survey questions. It also allows measurement in a context that is more natural than an interview. But observations can be distorted, too. Observers do not see or hear everything, and what they do see is filtered by their own senses and perspectives. Disagreements about crowd size among protestors, police, and journalists are notorious, even though there is a good method of estimating crowd size based on the "carrying capacity" of public spaces (McPhail and McCarthy 2004). When the goal is to observe behavior, measurement can be distorted because the presence of an observer may cause people to act differently than they would otherwise (Emerson 1983). I discuss these issues in more depth in Chapters 10 and 11; it is important to consider them whenever you read about observational measures.

Collecting Unobtrusive Measures

Unobtrusive measures allow us to collect data about individuals or groups without their direct knowledge or participation. In their classic book (now revised), Eugene Webb and his colleagues (2000) identified four types of unobtrusive measures: physical trace evidence, archives (available data), simple observation, and contrived observation (using hidden recording hardware or manipulation to elicit a response). We have already considered observational data, and we will consider available data (from "archives") in the next section, so I focus here on the other approaches suggested by Webb et al. (2000).

Physical Trace Evidence

The physical traces of past behavior are one type of unobtrusive measure that is most useful when the behavior of interest cannot be observed directly (perhaps because it is hidden or occurred in the past) and has not been recorded in a source of available data. To measure the prevalence of drinking in college dorms or fraternity houses, we might count the number of empty bottles of alcoholic beverages in the surrounding dumpsters. Student interest in the college courses they are taking might be measured by counting the number of times that books left on reserve as optional reading are checked out or by the number of class handouts left in trash barrels outside a lecture hall. Webb and his colleagues (2000:37) suggested measuring the interest in museum exhibits by the frequency with which tiles in front of the exhibits needed to be replaced. Social variables can also be measured by observing clothing, hair length, or people's reactions to such stimuli as dropped letters or jaywalkers.

Using Available Data

Government reports are rich and readily accessible sources of social science data. Organizations ranging from nonprofit service groups to private businesses also compile a wealth of figures that may be available to social scientists for some purposes. In addition, the data collected in many social science surveys are archived and made available for researchers who were not involved in the original survey project. For example, law enforcement and health statistics provide several community-level indicators of substance abuse (Gruenewald et al. 1997). Arrests for the sale and possession of drugs, drunk driving arrests, and liquor law violations (such as sales to minors) can usually be obtained on an annual basis and often quarterly from local police departments or state

crime information centers. Health-related indicators of substance abuse at the community level include single-vehicle fatal crashes, the rate of mortality from alcohol or drug abuse, and the use of alcohol and drug treatment services.

Indicators such as these cannot be compared across communities or over time without reviewing carefully how they were constructed. The level of alcohol in the blood that is legally required to establish intoxication can vary among communities, creating the appearance of different rates of substance abuse even though drinking and driving practices may be identical. Enforcement practices can vary among police jurisdictions and over time (Gruenewald et al. 1997:14).

We also cannot assume that available data are accurate, even when they appear to measure the concept in which we are interested in a way that is consistent across communities. "Official" counts of homeless persons have been notoriously unreliable because of the difficulty in locating homeless persons on the streets, and government agencies have, at times, resorted to "guesstimates" by service providers (Rossi 1989). Even available data for such seemingly straightforward measures as counts of organizations can contain a surprising amount of error. For example, a 1990 national church directory reported 128 churches in a midwestern U.S. county; an intensive search in that county in 1992 located 172 churches (Hadaway, Marler, and Chaves 1993:744). Perhaps 30% or 40% of death certificates identify incorrectly the cause of death (Altman 1998).

Government statistics that are generated through a central agency such as the U.S. Census Bureau are often of high quality, but caution is warranted when using official data collected by local levels of government. For example, the Uniform Crime Reports (UCR) program administered by the Federal Bureau of Investigation (FBI) imposes standard classification criteria, with explicit guidelines and regular training at the local level, but data are still inconsistent for many crimes. Officer-involved shootings are not reported reliably by many police districts (Wu 2021).

Focusing on a different level of aggregation may improve data quality because procedures for data collection may differ between cities, counties, states, and so on (Gruenewald et al. 1997:40–41). Only after factors such as legal standards, enforcement practices, and measurement procedures have been accounted for do comparisons between communities become credible.

Coding Content

Unobtrusive measures can also be created from diverse forms of media, such as newspaper archives or magazine articles, TV or radio talk shows, legal opinions, historical documents, personal letters, or e-mail messages. Content analysts may measure aspects of media such as the frequency of use of particular words or ideas or the consistency with which authors convey a particular message in their stories. An investigation of the drinking climate on campuses might include a count of the amount of space devoted to ads for alcoholic beverages in a sample of issues of the student newspaper.

Taking Pictures

Photographs record individual characteristics and social events, so they can become an important tool for investigating the social world. In recent years, photography has become a much more common part of the social world, as cameras embedded in smartphones and the use of websites and social media encourage taking and sharing photos (Tinkler 2013:15).

Figure 4.8 displays a photo of two Ukrainian women embroidering traditional Ukrainian Easter towels, but they are doing so while sitting on a park bench in Italy where they are domestic live-in workers. Olena Fedyuk (2012) included the picture in her analysis of photos exchanged by Ukrainian domestic workers in Italy with their family members who remained in Ukraine.

FIGURE 4.8 ■ Ukrainian Domestic Workers

Source: Oksana Pronyuk. From Fedyuk, Olena. 2012. "Images of Transnational Motherhood: The Role of Photographs in Measuring Time and Maintaining Connections Between Ukraine and Italy." *Journal of Ethnic and Migration Studies* 38:279–300.

Fedyuk concludes that pictures like these convey the reassuring message to their families that these women are still focused on their home country and family and are not engaged in Italian social life. Tragically, I don't need to tell you what very different message was conveyed by the horrid images coming out of Ukraine during the Russian invasion.

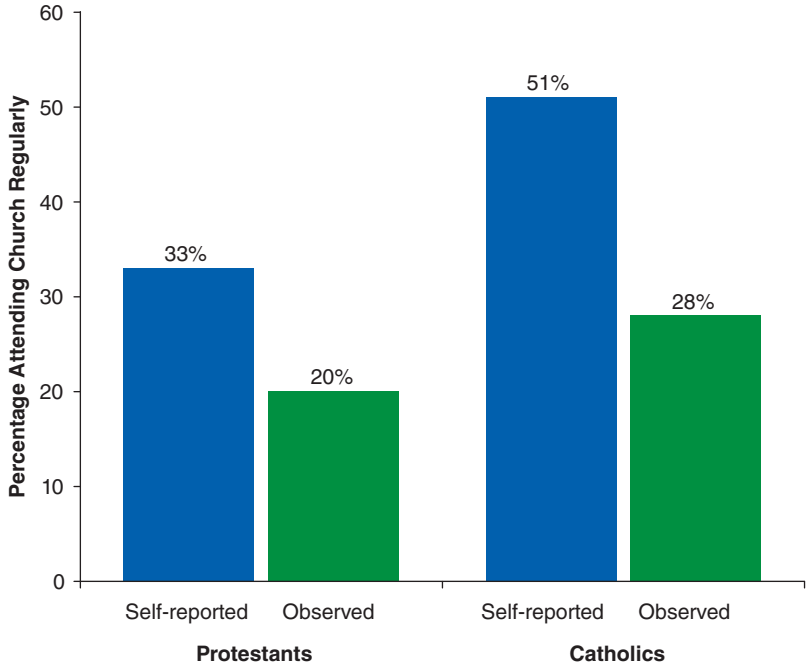
Combining Measurement Operations

Asking questions, making observations, using unobtrusive indicators, including available data or coding content, and taking pictures are interrelated measurement tools, each of which may include or be supplemented by the others. From people's answers to survey questions, the U.S. Census Bureau develops widely consulted census reports containing available data on people, firms, and geographic units in the United States. Data from employee surveys may be supplemented by information available in company records. Interviewers may record observations about those whom they question.

Questioning can be a particularly poor approach as the sole means for measuring socially desirable behaviors, such as voting or attending church, or that are socially stigmatized or illegal, such as abusing alcohol or drugs. Hadaway, Marler, and Mark (1993) demonstrated the tendency of people to answer questions in socially approved ways in a study of church attendance in the United States. More than 40% of adult Americans say in surveys that they attend church weekly—a percentage much higher than in Canada, Australia, or Europe. However, a comparison of observed church attendance with self-reported attendance suggested the actual rate of church attendance was much lower (see Figure 4.9).

Triangulation—the use of two or more different measures of the same variable—can thus strengthen measurement considerably (Brewer and Hunter 1989:17). When we achieve similar results with different measures of the same variable, particularly when they are based on such different methods as survey questions and field-based observations, we can be more confident in the validity of each measure. If results diverge with different measures, it may indicate that one or more of these measures are influenced by more measurement error than we can tolerate. Divergence between measures could also indicate they operationalize different concepts. For

FIGURE 4.9 ■ The Inadequacy of Self-Reports Regarding Socially Desirable Behavior: Observed Versus Self-Reported Church Attendance



Source: Data from Kirk C. Hadaway, Penny Long Marker, and Mark Chaves. 1993. "What the Polls Don't Show: A Closer Look at U.S. Church Attendance." *American Sociological Review* 58(6):741–752.

example, official crime statistics indicate only those crimes that are reported to and recorded by the police, but when surveys are used to measure the frequency of experiencing crimes, many "personal annoyances" are included as if they were crimes (Levine 1976).

LEVELS OF MEASUREMENT

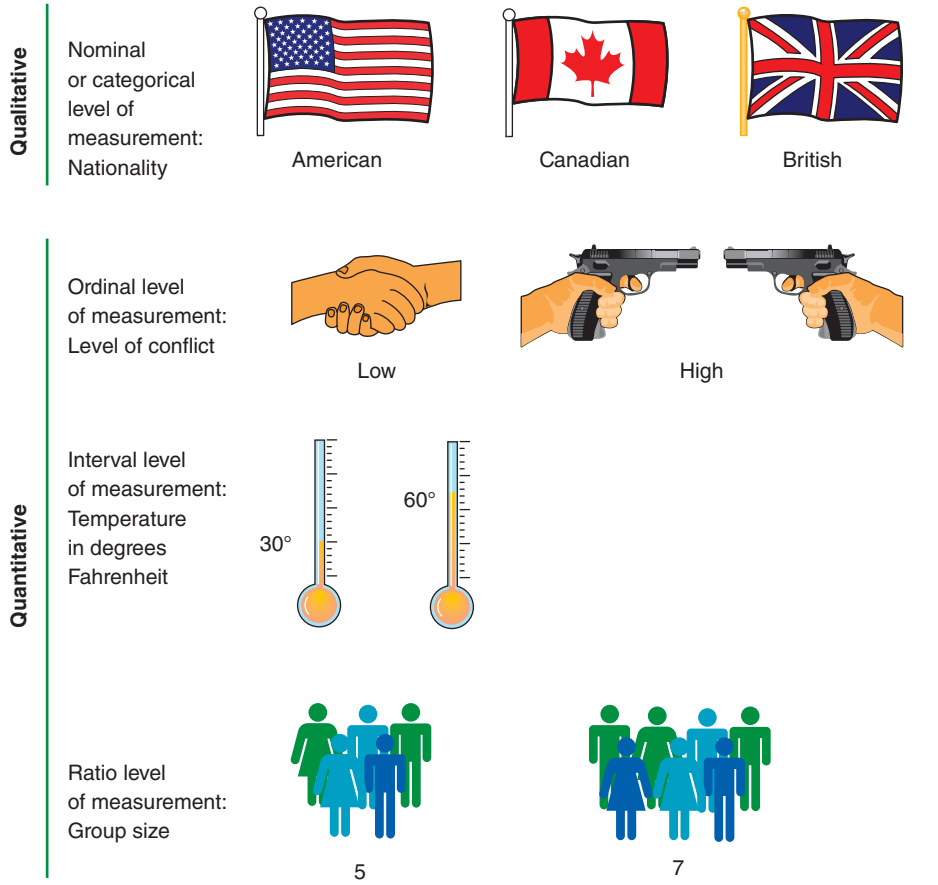
Can you name the variables represented in Figure 4.9? One variable is "religion"; it is represented by only two attributes, or categories, in Figure 4.9—Protestant and Catholic—but you know there are many others. You also know that one religion is not "more religion" than another; they are different in kind, not amount. The other variable represented in Figure 4.9 is "frequency of church attendance." Of course, frequencies *do* differ in amount. We can say that religion—a qualitative indicator—and frequency of church attendance—a quantitative indicator—differ in their levels of measurement.

When we know a variable's **level of measurement**, we understand more about how cases vary on that variable and so appreciate more fully what we have measured. Level of measurement also has important implications for the type of statistics that can be used with the variable, as you will learn in Chapter 9. There are four levels of measurement: (1) nominal, (2) ordinal, (3) interval, and (4) ratio. For most purposes, variables measured at the interval and ratio levels are treated in the same way, so I will sometimes refer to these two levels together as *interval–ratio.* Figure 4.10 depicts the differences among the four levels.

Nominal Level of Measurement

The **nominal level of measurement** identifies variables whose values have no mathematical interpretation; they vary in kind or quality but not in amount (they may also be called *categorical* or *qualitative variables*). It is conventional to refer to the values of nominal variables as *attributes* instead of values. The Monitoring the Future study (Jang et al. 2017:5) provides an example in the question asked about marital status. Response options were provided as 1 = married, 2 = engaged, 3 = separated/divorced, 4 = widowed, and 5 = single. The numbers represent the particular response options, but these numbers do not tell us anything about the difference between the types of marital status except that they are different. "Engaged" is not one unit less of *marital status* than "separated/divorced." Nationality, occupation, religious affiliation, and region of the country are also measured at the nominal level. A person may be Spanish or Portuguese, but one nationality does not represent more nationality than another—just a different nationality (see Figure 4.10). Of course, people may identify more strongly with one nationality than another, but these comparisons involve the variable, "strength of national identification," not *nationality per se.*

FIGURE 4.10 ■ Levels of Measurement



Although the attributes of categorical variables do not have a mathematical meaning, they must be assigned to cases with great care. The attributes we use to measure, or to categorize, cases must be mutually exclusive and exhaustive:

- A variable's attributes or values are mutually exclusive if every case can have only one attribute.
- A variable's attributes or values are exhaustive when every case can be classified into one of the categories.

When a variable's attributes are mutually exclusive and exhaustive, every case corresponds to one and only one attribute.

Ordinal Level of Measurement

The first of the three quantitative levels is the **ordinal level of measurement**. At this level, the numbers assigned to cases specify only the order of the cases, permitting *greater than* and *less than* distinctions. The properties of variables measured at the ordinal level are illustrated in Figure 4.10 by the contrast between the levels of conflict in two groups. The first group, symbolized by two people shaking hands, has a low level of conflict. The second group, symbolized by two persons pointing guns at each other, has a high level of conflict. To measure conflict, we would put the groups "in order" by assigning the number 1 to the low-conflict group and the number 2 to the high-conflict group. The numbers thus indicate only the relative position or order of the cases. Although low level of conflict is represented by the number 1, it is not one less unit of conflict than high level of conflict, which is represented by the number 2.

The Favorable Attitudes Toward Antisocial Behavior Scale measures attitudes toward antisocial behavior among high school students with a series of questions that each involves an ordinal distinction (see Figure 4.11). The response choices to each question range from "very wrong" to "not wrong at all." There's no real numerical quantity of "wrongness" that these distinctions reflect, but the idea is that a student who responds that it is "not wrong at all" to take a handgun to school has a more favorable attitude toward antisocial behavior than does a student who says it is "a little bit wrong"—which is in turn more favorable than those who respond "wrong" or "very wrong."

Like variables measured at the nominal level, variables measured at the ordinal level in this way classify cases in discrete categories and so are termed **discrete measures**.

A series of similar questions may be used instead of one question to measure the same concept. The set of questions in the Favorable Attitudes Toward Antisocial Behavior Scale in Figure 4.11 is a good example. In such a multi-item **index**, or *scale*, numbers are assigned to reflect the order of the responses (such as 1 for "very wrong," 2 for "wrong," 3 for "a little bit wrong," and 4 for "not wrong at all"); these responses are then summed or averaged to create the index score. One person's responses to the five questions in Figure 4.11 could thus range from 5 (meaning they said each behavior is "very wrong") to 20 (meaning they said each behavior is "not wrong at all"). However, even though these are numeric scores, they still reflect an ordinal level of measurement because the responses they are based on involve only ordinal distinctions.

FIGURE 4.11 ■ Example of Ordinal Measures: Favorable Attitudes Toward Antisocial Behavior Scale

1. How wrong do you think it is for someone your age to take a handgun to school?
   Very wrong | Wrong | A little bit wrong | Not wrong at all
2. How wrong do you think it is for someone your age to steal anything worth more than $5?
   Very wrong | Wrong | A little bit wrong | Not wrong at all
3. How wrong do you think it is for someone your age to pick a fight with someone?
   Very wrong | Wrong | A little bit wrong | Not wrong at all
4. How wrong do you think it is for someone your age to attack someone with the idea of seriously hurting them?
   Very wrong | Wrong | A little bit wrong | Not wrong at all
5. How wrong do you think it is for someone your age to stay away from school all day when their parents think they are at school?
   Very wrong | Wrong | A little bit wrong | Not wrong at all

Sources: Lewis, Chandra, Gwen Hyatt, Keith Lafortune, and Jennifer Lembach. 2010. *History of the Use of Risk and Protective Factors in Washington State's Healthy Youth Survey.* Portland, OR: RMC Research Corporation. Retrieved May 11, 2014 (https://www.askhys.net/library/Old/RPHistory.pdf), p. 26. See also Arthur, Michael W., John S. Briney, J. David Hawkins, Robert D. Abbott, Blair L. Brooke-Weiss, and Richard F. Catalano. 2007. "Measuring Risk and Protection in Communities Using the Communities That Care Youth Survey." *Evaluation and Program Planning* 30:197–211.

Interval Level of Measurement

The numbers indicating the values of a variable at the **interval level of measurement** represent fixed measurement units but have no absolute, or fixed, zero point. This level of measurement is represented in Figure 4.10 by the difference between two Fahrenheit temperatures. Although 60 degrees is 30 degrees hotter than 30 degrees, 60, in this case, is not twice as hot as 30. Why not? Because "heat" does not begin at 0 degrees on the Fahrenheit scale.

An interval-level measure is created by a scale that has fixed measurement units but no absolute, or fixed, zero point. The numbers can, therefore, be added and subtracted, but ratios are not meaningful. Again, the values must be mutually exclusive and exhaustive. There are few true interval-level measures in the social sciences, but cross-disciplinary investigations such as those that examine the linkage between climate and social life or historical change in social organization may involve interval measures such as temperature or calendar year.

Many social scientists use indexes created by combining responses to a series of variables measured at the ordinal level as if they were interval-level measures. Figure 4.11 provided one example of such a multi-item index. Scores on scales that are standardized in a distribution can also be treated as measured at the interval level. For example, the so-called intelligence quotient, or IQ score, is standardized so that a score of 100 indicates a person is in the middle of the distribution, and each score above and below 100 indicates fixed points on the distribution of IQ scores in the population. A score of 114 would represent a score higher than that of 84% of the population.

Ratio Level of Measurement

The numbers indicating the values of a variable at the **ratio level of measurement** represent fixed measuring units and an absolute zero point (zero means absolutely no amount

of whatever the variable indicates). For example, the following question was used on the National Minority SA/HIV Prevention Initiative Youth Questionnaire to measure the number of days during the past 30 days that the respondent drank at least one alcoholic beverage (Figure 4.12). We can easily calculate the number of days that separate any response from any other response (except for the value for "don't know," which must be treated as missing data).

Figure 4.10 also displays an example of a variable measured at the ratio level. The number of people in the first group is five, and the number in the second group is seven. The ratio of the two groups' sizes is then 1.4, a number that mirrors the relationship between the sizes of the groups. Note there does not actually have to be any group with a size of zero; what is important is that the numbering scheme begins at an absolute zero—in this case, the absence of any people.

For most statistical analyses in social science research, the interval and ratio levels of measurement can be treated as equivalent. In addition to having numerical values, both the interval and ratio levels also involve continuous measures: The numbers indicating the values of variables are points on a continuum, not discrete categories. But despite these similarities, there is an important difference between variables measured at the interval and ratio levels. On a ratio scale, 10 is 2 points higher than 8 and is also 2 *times* greater than 5—the numbers can be compared in a ratio. Ratio numbers can be added and subtracted, and because the numbers begin at an absolute zero point, they can be multiplied and divided (so ratios can be formed between the numbers). For example, people's ages can be represented by values ranging from 0 years (or some fraction of a year) to 120 or more. A person who is 30 years old is 15 years older than someone who is 15 years old (30 – 15 = 15) and is twice as old as that person (30/15 = 2). Of course, the numbers also are mutually exclusive and exhaustive, so that every case can be assigned one and only one value.

FIGURE 4.12 ■ Assessment of Alcoholic Beverage Consumption

During the past 30 days, on how many days did you drink one or more drinks of an alcoholic beverage?

¨ 0 days	¨ 12 days	¨ 24 days
¨ 1 day	¨ 13 days	¨ 25 days
¨ 2 days	¨ 14 days	¨ 26 days
¨ 3 days	¨ 15 days	¨ 27 days
¨ 4 days	¨ 16 days	¨ 28 days
¨ 5 days	¨ 17 days	¨ 29 days
¨ 6 days	¨ 18 days	¨ 30 days
¨ 7 days	¨ 19 days	¨ Don't know or can't say
¨ 8 days	¨ 20 days	
¨ 9 days	¨ 21 days	
¨ 10 days	¨ 22 days	
¨ 11 days	¨ 23 days	

The Special Case of Dichotomies

Dichotomies, variables having only two values, are a special case from the standpoint of levels of measurement. The values or attributes of a variable such as "region of origin" clearly vary in kind or quality but not in amount. Thus, the variable is categorical—measured at the nominal level. Yet if region is dichotomized into North and South, we can also think of the variable as indicating the presence of the attribute *northern* (or *southern*) or not. Viewed in this way, there is an inherent order: Someone born in the South has more of the southern attribute (it is present) than a northerner (the attribute is not present). It's also possible to think of a **dichotomy** as representing an interval level of measurement because there is an equal interval between the two attributes. So what do you answer to the test question, "What is the level of measurement of *region of origin?*" "Nominal," of course, but you'll find that when a statistical procedure requires that variables be quantitative, a dichotomy can be perfectly acceptable.

Comparison of Levels of Measurement

Table 4.1 summarizes the types of comparisons that can be made with different levels of measurement, as well as the mathematical operations that are legitimate. Each higher level of measurement allows a more precise mathematical comparison to be made between the value measured at that level compared with those measured at lower levels. However, each comparison between cases measured at lower levels can also be made about cases measured at the higher levels. Thus, all four levels of measurement allow researchers to assign different values to different cases. All three quantitative measures allow researchers to rank cases in order.

Researchers choose the levels of measurement in the process of operationalizing variables; the level of measurement is not inherent in the variable itself. Many variables can be measured at different levels, with different procedures. For example, the Core Alcohol and Drug Survey (Core Institute 1994) identifies binge drinking by asking students, "Think back over the last two weeks. How many times have you had five or more drinks at a sitting?" You might be ready to classify this as a ratio-level measure, but this would be true only if responses are recorded as the actual number of "times." Instead, the Core Survey treats this as a closed-ended question, and students are asked to indicate their answer by checking "None," "Once," "Twice," "3 to 5 times," "6 to 9 times," or "10 or more times." Use of these categories makes the level of measurement ordinal because the distance between any two cases cannot be clearly determined. A student

TABLE 4.1 ■ Properties of Levels of Measurement

Level of Measurement	Possible Statements: Age of Two Individuals: Sara (24) and Bill (12)	Quality of Level	Appropriate Math Operations
Nominal	Adults (over 18) Adolescent (under 18)	Different from	= (?)
Ordinal	Sara is older than Bill; Bill is younger than Sara.	Greater than	> (<)
Interval	Sara is 12 years older than Bill; Bill is 12 years younger than Sara	A year is a year is a year	+ (-)
Ratio	Bill is ½ Sara's age or Sara is twice Bill's age	Absolute zero	× (÷)

Source: Adapted from material provided by Tajuana D. Massie, assistant professor, social sciences, South Carolina State University.

with a response in the "6 to 9 times" category could have binged just one more time or six more times than a student who responded "3 to 5 times." You just can't tell.

It is usually a good idea to try to measure variables at the highest level of measurement possible. The more information available, the more ways we have to compare cases. Be aware, however, that other considerations may suggest *not* measuring at the highest level possible. For example, many people are very reluctant to report their exact incomes, even in anonymous questionnaires. As a result, asking respondents to report their income in categories (such as less than $10,000, $10,000–$19,999, $20,000–$29,999) will result in more responses and thus more valid data than will asking respondents for their income in dollars.

As I've already mentioned, researchers often treat variables measured at the interval and ratio levels as comparable. They then refer to this as the interval–ratio level of measurement. You will learn in Chapter 9 that different statistical procedures are used for variables with fixed measurement units, but it usually doesn't matter whether there is an absolute zero point.

EVALUATING MEASURES

Do the operations developed to measure our variables actually do so—are they valid? If we have carefully constructed our procedures, we should be on the right track. But we cannot have much confidence in a measure until we have empirically evaluated its validity. What good is our measure if it doesn't measure what we think it does? If our measurement procedure is invalid, we might as well go back to the starting block and try again. As part of evaluating the validity of our measures, we must also evaluate their reliability because reliability (consistency) is a prerequisite for measurement validity.

Measurement Validity

In Chapter 2, you learned that measurement validity refers to the extent to which measures indicate what they are intended to measure. More technically, a valid measure of a concept is one that is closely related to other apparently valid measures of the concept and to the known or supposed correlates of that concept, but that is not related to measures of unrelated concepts, irrespective of the methods used for the other different measures (Brewer and Hunter 1989:134).

When a measure "misses the mark"—when it is not valid—our measurement procedure has been affected by measurement error. Measurement error has two sources:

1. **Idiosyncratic (or "random") errors** are errors that affect individuals or cases in unique ways that are unlikely to be repeated in just the same way (Viswanathan 2005:289). Individuals make idiosyncratic errors when they don't understand a question, when some unique feelings are triggered by the wording of a question, or when they are feeling out of sorts because of some recent events. Idiosyncratic errors may arise in observational research when the observer is distracted or misperceives an event. Some idiosyncratic error is unavoidable with any measurement procedure, although it is important to reduce such error as much as possible. However, because such errors are idiosyncratic, or random, they are as likely to be above as below the true value of the measure; this means that idiosyncratic errors should not bias the measure in one direction.

2. **Systematic errors** occur when responses are affected consistently by factors that are not what the instrument is intended to measure. For example, individuals who like

to please others by giving socially desirable responses may tend to say they "agree" with statements, simply because they try to avoid saying they "disagree" with anyone. Systematic errors may also arise when the same measure is used across cultures that differ in their understanding of the concepts underlying the measures (Church 2010:152–153). In addition, **unbalanced response choices** may lead most respondents to give positive rather than negative responses. For example, if respondents are asked the question with the unbalanced response choices in Figure 4.13, they are more likely to respond that gun ownership is wrong than if they are asked the question with the **balanced response choices** (Viswanathan 2005:142–148).

FIGURE 4.13 ■ Balanced and Unbalanced Response Choices

Unbalanced response choices

How wrong do you think it is for someone who is not a hunter to own a gun?

- Very wrong
- Wrong
- A little bit wrong
- Not wrong at all

Balanced response choices

Some people think it is wrong for someone who is not a hunter to own a gun, and some people think it is a good idea to own a gun. Do you think it is very wrong, wrong, neither right nor wrong, right, or very right for someone who is not a hunter to own a gun?

- Very wrong
- Wrong
- Neither right nor wrong
- Right
- Very right

Systematic errors can do much more damage to measurement validity than idiosyncratic errors can because they will lead to the average value of the indicator being higher or lower than the phenomenon that it is measuring. For example, if a political poll uses questions that generate agreement bias, a politician may believe that voters agree with their position much more than they actually do.

The social scientist must try to reduce measurement errors and then to evaluate the extent to which the resulting measures are valid. The extent to which measurement validity has been achieved can be assessed with four different approaches: (1) face validation, (2) content validation, (3) criterion validation, and (4) construct validation. The methods of criterion and construct validation also include subtypes.

Face Validity

Researchers apply the term **face validity** to the confidence gained from careful review of a measure to see if it seems appropriate "on its face." More precisely, we can say that a measure is face valid if it obviously pertains to the meaning of the concept being measured more than to other concepts (Brewer and Hunter 1989:131). For example, a count of the number of drinks people had consumed in the past week would be a face-valid measure of their alcohol consumption.

Although every measure should be inspected in this way, face validation in itself does not provide convincing evidence of measurement validity (DeVellis 2017:100–102). The question "How much beer or wine did you have to drink last week?" looks valid on its face as a measure of frequency of drinking, but people who drink heavily tend to underreport the amount they drink. So the question would be an invalid measure, at least in a study of heavy drinkers.

Content Validity

Content validity establishes that the measure covers the full range of the concept's meaning. To determine that range of meaning, the researcher may solicit the opinions of experts and review literature that identifies the different aspects, or dimensions, of the concept.

An example of a measure that covers a wide range of meaning is the Michigan Alcoholism Screening Test (MAST). The MAST includes 24 questions representing the following subscales: recognition of alcohol problems by self and others; legal, social, and work problems; help seeking; marital and family difficulties; and liver pathology (Skinner and Sheu 1982). Many experts familiar with the direct consequences of substance abuse agree that these dimensions capture the full range of possibilities. Thus, the MAST is believed to be valid from the standpoint of content validity.

Criterion Validity

Criterion validity is established when the scores obtained on one measure can be compared accurately with those obtained with a more direct or already validated measure of the same phenomenon (the criterion). A measure of blood alcohol concentration or a urine test could serve as the criterion for validating a self-report measure of drinking, as long as the questions we ask about drinking refer to the same time period. Chemical analysis of hair samples can reveal unacknowledged drug use (Mieczkowski 1997). Friends' or relatives' observations of a person's substance use also could serve, in some limited circumstances, as a criterion for validating self-report substance use measures.

The criterion that researchers select can be measured either at the same time as the variable to be validated or after that time. **Concurrent validity** exists when a measure yields scores that are closely related to scores on a criterion measured at the same time. A store might validate a question-based test of sales ability by administering it to sales personnel who are already employed and then comparing their test scores with their sales performance. **Predictive validity** is the ability of a measure to predict scores on a criterion measured in the future. For example, a store might administer a test of sales ability to new sales personnel and then validate the measure by comparing these test scores with the criterion—the subsequent sales performance of the new personnel.

An attempt at criterion validation is well worth the effort because it greatly increases confidence that the standard is measuring what was intended. However, for many concepts of interest to social scientists, no other variable can reasonably be considered a criterion. If we are measuring feelings or beliefs or other subjective states, such as feelings of "alienation," what *direct* indicator could serve as a criterion? Even with variables for which a reasonable criterion exists, the researcher may not be able to gain access to the criterion—as would be the case with a tax return or employer document that we might wish we could use as a criterion for self-reported income.

Construct Validity

Measurement validity can also be established by showing that a measure is related to a variety of other measures as specified in a theory or in prior research. This validation approach,

known as **construct validity**, is commonly used in social research when no clear criterion exists for validation purposes. For example, in an analysis of support for the Black Lives Matter (BLM) movement using data from the 2016 American National Election Studies, Kevin Drakulich and collegues (2021) found that level of support for BLM tended to decline among those who felt more warmth toward the police, had more "cold" feelings toward Black persons, and reported higher levels of racial resentment. These associations were predicted by theory and prior research and could be seen as supporting the construct validity of the question measuring support for BLM.

Two other approaches to construct validation are convergent validation and discriminant validation. **Convergent validity** is achieved when one measure of a concept is associated with different types of measures of the same concept (this relies on the same type of logic as measurement triangulation). **Discriminant validity** is a complementary approach to construct validation. In this approach, scores on the measure to be validated are compared with scores on measures of different but related concepts. Discriminant validity is achieved if the measure to be validated is not associated strongly with the measures of different concepts.

Measurement Reliability

Reliability means that a measurement procedure yields consistent scores when the phenomenon being measured is not changing (or that the measured scores change in direct correspondence to actual changes in the phenomenon). If a measure is reliable, it is affected less by random error, or chance variation, than if it is unreliable. Reliability is a prerequisite for measurement validity: We cannot really measure a phenomenon adequately if the measure we are using gives inconsistent results. Because it is usually easier to assess reliability than validity, you are more likely to see an evaluation of measurement reliability in a research report than an evaluation of measurement validity.

Problems in reliability can occur when inconsistent measurements are obtained after the same phenomenon is measured multiple times, with multiple indicators, or by multiple observers.

Multiple Times: Test–Retest and Alternate Forms

When researchers measure a phenomenon that does not change between two points separated by an interval of time, the degree to which the two measurements are related to each other is the **test–retest reliability** of the measure. If you take a test of your math ability and then retake the test two months later, the test is performing reliably if you receive a similar score both times—presuming that nothing happened during the two months to change your math ability. Of course, if events between the test and the retest have changed the variable being measured (maybe due to completing two months of a good course math course), then the difference between the test and retest scores should reflect that change.

When ratings by an observer rather than ratings by the subjects themselves are being assessed at two or more points in time, test–retest reliability is termed **intrarater (or intraobserver) reliability**. If an observer's ratings of individuals' drinking behavior in bars are similar at two or more points in time and the behavior has not changed, the observer's ratings of drinking behavior are reliable.

Researchers test **alternate-forms reliability** by comparing the subjects' answers with slightly different versions of the survey questions (Litwin 1995:13–21). A researcher may reverse the order of the response choices in an index or modify the question wording in minor ways and then readminister that index to the subjects. If the two sets of responses are not too different, alternate-forms reliability is established.

Multiple Indicators: Interitem and Split-Half

When researchers use multiple items to measure a single concept, they must be concerned with **interitem reliability** (or *internal consistency*). For example, if we are to have confidence that a set of questions (such as those in Table 4.2) reliably measures depression, the answers to these questions should be highly associated with one another. The stronger the association between the individual items and the more items included, the higher the reliability of the index. **Cronbach's alpha** is a **reliability measure** commonly used to measure interitem reliability. Of course, interitem reliability cannot be computed if only one question is used to measure a concept. For this reason, it is much better to use a multi-item index to measure an important concept (Viswanathan 2005:298–299). Donald Hawkins, Paul Amato, and Valarie King (2007:1007) used the "CES-D" index to measure depression in their study of adolescent well-being and obtained a high level of interitem reliability (see Table 4.2).

A related test of reliability is the **split-half reliability** approach. After a set of survey questions intended to form an index of one concept is administered, the researcher divides the questions into half by distinguishing even- and odd-numbered questions, flipping a coin, or using some other random procedure. Scores are then computed for these two sets of questions. The researchers then compare the scores for the two halves and check the relation between the subjects' scores on them. If scores on the two halves are similar and highly related to each other (so that people who score high on one half also score high on the other half, etc.), then the measure's split-half reliability is established.

TABLE 4.2 ■ Examples of Indexes: Short Form of the Center for Epidemiologic Studies (CES-D) and "Negative Outlook" Index

At any time during the past week . . . (Circle one response on each line)	Never	Some of the Time	Most of the Time
a. Was your appetite so poor that you did not feel like eating?	1	2	3
b. Did you feel so tired and worn out that you could not enjoy anything?	1	2	3
c. Did you feel depressed?	1	2	3
d. Did you feel unhappy about the way your life is going?	1	2	3
e. Did you feel discouraged and worried about your future?	1	2	3
f. Did you feel lonely?	1	2	3

Source: Hawkins, Daniel N., Paul R. Amato, and Valarie King. 2007. "Nonresident Father Involvement and Adolescent Well-Being: Father Effects or Child Effects?" *American Sociological Review* 72:990.

Multiple Observers: Interobserver and Intercoder

When researchers use more than one observer to rate the same people, events, or places, **interobserver reliability** is their goal. If observers are using the same instrument to rate the same thing, their ratings should be very similar. If they are similar, we can have much more confidence that the ratings reflect the phenomenon being assessed rather than the orientations of the observers.

Assessing interobserver reliability is most important when the rating task is complex. Consider a measure of neighborhood disorder, the African American Health Neighborhood Assessment Scale (AAH NAS), which is shown in Figure 4.14. The rating task seems straightforward, with clear descriptions of the block characteristics that are rated to produce an overall neighborhood score. However, the judgments the rater must make while using this scale are complex. They are affected by a wide range of subject characteristics, attitudes, and behaviors as well as by the rater's reactions. As a result, achieving reliable ratings required careful training of the raters.

FIGURE 4.14 ■ The Challenge of Interobserver Reliability: African American Health Seven-Item Neighborhood Assessment Scale (AAH NAS)

The first 3 questions refer to the full block and street on which the respondent lives.

1. Volume of traffic
   0. No Traffic
   1. Light (occasional cars)
   2. Moderate
   3. Heavy (steady stream of cars)
2. Condition of the street
   0. Under construction
   1. Very poor (many sizeable cracks, potholes, or broken curbs)
   2. Fair
   3. Moderately good (no sizeable cracks, potholes, or broken curbs)
   4. Very good
3. How noisy is the street?
   0. Very quiet—easy to hear almost anything
   1. Fairly quiet—can hear people walking by talking, though you may not understand them
   2. Somewhat noisy—voices are not audible unless very near
   3. Very noisy—difficult to hear a person talking near to you

Items 4 through 7 are answered based on observations of the side of the street on the block where the respondent lives (block face).

4. Are empty beer or liquor bottles in street, yard, or alley present?
   1. Yes
   0. No
5. Are there cigarette or cigar butts or discarded cigarette packages on the sidewalk or in the gutters?
   1. Yes
   0. No
6. Is there garbage, litter, or broken glass in the street or on the sidewalks?
   0. None
   1. Light (some visible)
   2. Moderate
   3. Heavy (visible along most or all of street)
7. In general, how would you rate the condition of most of the residential units in the block face?
   0. Very well kept/good condition—attractive for its type
   1. Moderately well kept condition
   2. Fair condition (peeling paint, needs repair)
   3. Poor/Badly deteriorated condition

Source: Andresen E., T. K. Malmstrom, F. D. Wolinsky, M. Schootman, J. P. Miller, and D. K. Miller. "Rating Neighborhoods for Older Adult Health: Results from the African American Health Study." *BMC Public Health* 2008, 8:35. doi:10.1186/1471-2458-8-35.

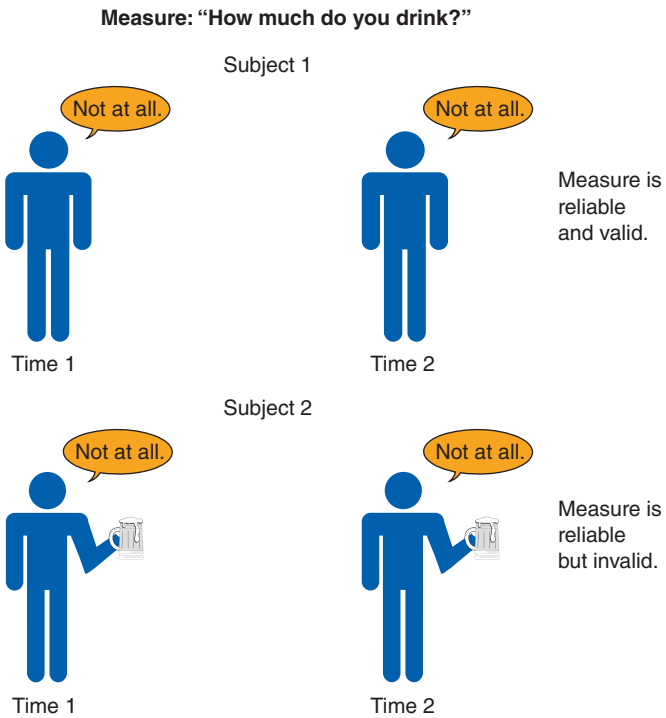
It is also important to establish an adequate level of **intercoder reliability** when data are transferred from their original form, whether observations or interviews, into structured codes or simply into a data entry program or spreadsheet. There can be weak links in data processing, so the consistency of coders should be tested.

Ways to Improve Reliability and Validity

Whatever the concept or the validation method used, no measure is without some error, nor can we expect it to be valid for all times and places. For example, the reliability and validity of self-report measures of substance abuse vary with factors such as whether the respondents are sober or intoxicated at the time of the interview, whether the measure refers to recent or lifetime abuse, and whether the respondents see their responses as affecting their chances of receiving housing, treatment, or some other desired outcome (Babor, Stephens, and Marlatt 1987). We should always be on the lookout for ways in which we can improve the reliability and validity of the measures we use.

Remember a reliable measure is not necessarily a valid measure, as Figure 4.15 illustrates. This discrepancy is a common flaw of self-report measures of substance abuse. Most respondents answer the multiple questions in self-report indexes of substance abuse in a consistent way, so the indexes are reliable. However, a number of respondents will not admit to drinking, even though they drink a lot. Their answers to the questions are consistent, but they are consistently misleading. As a result, some indexes based on self-report are reliable but invalid. Such indexes are not useful and should be improved or discarded. Unfortunately, many measures are judged to be worthwhile on the basis of only a reliability test, without assessing their validity.

FIGURE 4.15 ■ The Difference Between Reliability and Validity: Drinking Behavior



The reliability and validity of measures in any study must be tested after the fact to assess the quality of the information obtained. But then, if it turns out that a measure cannot be considered reliable and valid, little can be done to save the study. Hence, it is supremely important to select, in the first place, measures that are likely to be reliable and valid. Don't just choose the first measure you find or can think of: Consider the different strengths of different measures and their appropriateness to your study. Conduct a pretest in which you use the measure with a small sample and check its reliability. Provide careful training to ensure a consistent approach if interviewers or observers will administer the measures. In most cases, however, the best strategy is to use measures that have been used before and whose reliability and validity have been established in other contexts. But the selection of tried-and-true measures still does not absolve researchers from the responsibility of testing the reliability and validity of the measure in their own studies.

If the research focuses on previously unmeasured concepts, new measures will have to be devised. Researchers can use one of three strategies to improve the likelihood that new question-based measures will be reliable and valid (Fowler 1995):

> *Engage potential respondents in group discussions about the questions to be included in the survey.* This strategy allows researchers to check for consistent understanding of terms and to hear the range of events or experiences that people will report.
>
> *Conduct cognitive interviews.* Ask people a test question, then probe with follow-up questions about how they understood the question and what their answer meant.
>
> *Audiotape test interviews during the pretest phase of a survey.* The researchers then review these audiotapes and systematically code them to identify problems in question wording or delivery. (pp. 104–129)

In these ways, qualitative methods help improve the validity of the fixed-response questions used in quantitative surveys. I will discuss these methods in more detail in Chapter 8.

CONCLUSIONS

Remember always that measurement validity is a necessary foundation for social research. Gathering data without careful conceptualization or conscientious efforts to operationalize key concepts often is a wasted effort.

The difficulties of achieving valid measurement vary with the concept being operationalized and the circumstances of the study. The examples in this chapter of difficulties in achieving valid measures of substance abuse should sensitize you to the need for caution. However, don't let these difficulties discourage you: Substance abuse is a relatively difficult concept to operationalize because it involves behavior that is socially stigmatized and often illegal. Most other concepts in social research present fewer difficulties. But even substance abuse can be measured adequately with a proper research design.

Planning is the key to achieving valid measurement in your own research; careful evaluation is the key to sound decisions about the validity of measures in others' research. Statistical tests can help determine whether a given measure is valid after data have been collected, but if it appears after the fact that a measure is invalid, little can be done to correct the situation. If you cannot tell how key concepts were operationalized when you read a research report, don't trust the

findings. And if a researcher does not indicate the results of tests used to establish the reliability and validity of key measures, remain skeptical.

KEY TERMS

Alternate-forms reliability (p. 140)
Balanced response choices (p. 138)
Closed-ended (fixed-choice) question (p. 126)
Concept (p. 113)
Conceptualization (p. 113)
Concurrent validity (p. 139)
Constant (p. 119)
Construct validity (p. 140)
Content validity (p. 139)
Continuous measure
Convergent validity (p. 140)
Criterion validity (p. 139)
Cronbach's alpha (p. 141)
Dichotomy (p. 136)
Discrete measure
Discriminant validity (p. 140)
Exhaustive (p. 126)
Face validity (p. 138)
Idiosyncratic (or "random") errors (p. 137)
Index (p. 133)
Indicator
Intercoder reliability (p. 143)
Interitem reliability (p. 141)
Interobserver reliability (p. 141)
Interval level of measurement (p. 134)
Interval–ratio level of measurement
Intrarater (or intraobserver) reliability (p. 140)
Level of measurement (p. 131)
Measurement (p. 123)
Mutually exclusive (p. 126)
Nominal level of measurement (p. 132)
Open-ended question
Operationalization (p. 118)
Ordinal level of measurement (p. 133)
Predictive validity (p. 139)
Ratio level of measurement (p. 134)
Reliability (p. 140)
Reliability measures
Split-half reliability (p. 141)
Systematic errors (p. 137)
Test–retest reliability (p. 140)
Triangulation (p. 130)
Unbalanced response choices (p. 138)
Unobtrusive measure

HIGHLIGHTS

LO 4.1 Explain why the conceptualization of terms we use in our research is important.

- Conceptualization is the process of specifying what we mean by a term.
- Conceptualization is important because we cannot be certain that all readers will share a particular definition or that the current meaning of the concept is the same as it was when previous research was published.
- It is especially important to clearly define any concepts that are abstract or unfamiliar. When we refer to concepts such as *social control, social health, caste*, or *anomie,* we cannot count on others knowing exactly what we mean. Even experts may disagree about the meaning of frequently used concepts if they have based their conceptualizations on different theories.
- In deductive research, conceptualization helps translate portions of an abstract theory into specific variables that can be used in testable hypotheses.
- In inductive research, conceptualization is an important part of the process used to make sense of related observations.

LO 4.2 Discuss the relationship between conceptualization and operationalization.

- Operationalization involves connecting concepts to measurement operations. Once we have clearly defined the concepts we use, we develop specific procedures for indicating variation in the variables related to these concepts.
- It is particularly important to specify the indicators to be used to operationalize a new and emerging concept like perceived threat from immigrants.
- Qualitative research projects usually take an inductive approach to the process of conceptualization. In an inductive approach, concepts emerge from the process of thinking about what has been observed.
- Quantitative projects usually take a deductive approach in which we develop concepts based on theory and then decide what should be observed to indicate that concept.

LO 4.3 Illustrate key forms of measurement, including surveys, observations, unobtrusive measures, and published data.

- Closed-ended (fixed-choice) question: A survey question that provides preformatted response choices for the respondent to circle or check.
- Response choices should be mutually exclusive and exhaustive, so that every respondent can find one and only one choice that applies to him or her (unless the question is of the "Check all that apply" format).
- Most surveys of large numbers of people contain primarily fixed-choice questions, which are easy to process with computers and analyze with statistics.
- With fixed-choice questions, respondents are also more likely to answer the questions that the researcher really wants them to answer.
- Including response choices reduces ambiguity and makes it easier for respondents to answer.
- However, fixed-response choices can obscure what people really think if the choices do not match the range of possible responses to the question.
- Open-ended question: A survey question to which the respondent replies in his or her own words, either by writing or by talking.
- In qualitative research, open-ended questions are often used to explore the meaning respondents give to abstract concepts.
- Open-ended questions are preferable when the range of responses cannot adequately be anticipated—namely, questions that have not previously been used in surveys and questions that are asked of new groups—or when there is simply a long list of possible responses.
- Open-ended questions can also lessen confusion about the meaning of responses involving complex concepts.

LO 4.4 Produce examples of the four levels of measurement for the concept.

- Level of measurement is the mathematical precision with which the values of a variable can be expressed.
- There are four levels of measurement: (1) nominal, (2) ordinal, (3) interval, and (4) ratio.
- Nominal level of measurement: variables whose values have no mathematical interpretation; they vary in kind or quality but not in amount.

- Ordinal level of measurement: a measurement of a variable in which the numbers indicating a variable's values specify only the order of the cases, permitting *greater than* and *less than* distinctions.
- Interval level of measurement: a measurement of a variable in which the numbers indicating a variable's values represent fixed measurement units but have no absolute, or fixed, zero point.
- Ratio level of measurement: a measurement of a variable in which the numbers indicating a variable's values represent fixed measuring units and an absolute zero point.

LO 4.5 Compare the different approaches to measurement validity and reliability and recommend ways to improve validity and reliability.

- The extent to which measurement validity has been achieved can be assessed with four different approaches: (1) face validation, (2) content validation, (3) criterion validation, and (4) construct validation.
  - **Face validity** is the confidence gained from careful review of a measure to see if it seems appropriate "on its face." Although every measure should be inspected in this way, face validation does not provide convincing evidence of measurement validity.
  - **Content validity** establishes that the measure covers the full range of the concept's meaning. To determine that range of meaning, the researcher may solicit the opinions of experts and review literature that identifies the different aspects, or dimensions, of the concept.
  - **Criterion validity** is established when the scores obtained on one measure can be compared accurately with those obtained with a more direct or already validated measure of the same phenomenon (the criterion).

    The criterion that researchers select can be measured either at the same time as the variable to be validated or after that time.
  - **Concurrent validity** exists when a measure yields scores that are closely related to scores on a criterion measured at the same time.
  - **Predictive validity** is the ability of a measure to predict scores on a criterion measured in the future.
  - **Construct validity** is established when a measure is related to a variety of other measures as specified in a theory. This is commonly used in social research when no clear criterion exists for validation purposes.
  - **Convergent validity** is achieved when one measure of a concept is associated with different types of measures of the same concept (this relies on the same type of logic as measurement triangulation).
  - **Discriminant validity** is a complementary approach to construct validation. In this approach, scores on the measure to be validated are compared with scores on measures of different but related concepts. Discriminant validity is achieved if the measure to be validated is not associated strongly with the measures of different concepts.
- Reliability can be assessed in three different ways: through assessments being performed multiple times, with multiple indicators, and with multiple observers.
- Multiple times: A researcher can pursue test–retest reliability , a measurement that shows whether measures of a phenomenon at two points in time are highly correlated--if the phenomenon has not changed or has changed only as much

as the measures have changed. This form of reliability is also called intrarater or intraobserver reliability. A researcher can also assess alternate-forms reliability, a procedure for testing the reliability of responses to survey questions in which subjects' answers are compared after the subjects have been asked slightly different versions of the questions or when randomly selected halves of the sample have been administered slightly different versions of the questions.

- Multiple indicators: When researchers use multiple items to measure a single concept, they are concerned with interitem reliability (or *internal consistency*). The stronger the association between the individual items and the more items included, the higher the reliability of the index. Cronbach's alpha is a reliability measure commonly used to measure interitem reliability. A researcher may also assess split-half reliability, which is reliability achieved when responses to the same questions by two randomly selected halves of a sample are about the same.
- Multiple observers: When researchers use more than one observer to rate the same people, events, or places, interobserver reliability is their goal. It is also important to establish an adequate level of intercoder reliability when data are transferred from their original form, whether observations or interviews, into structured codes or simply into a data entry program or spreadsheet.
- Researchers can use one of three strategies to improve the likelihood that new question-based measures will be reliable and valid (Fowler 1995):
  - *Engage potential respondents in group discussions about the questions to be included in the survey.* This strategy allows researchers to check for a consistent understanding of terms and to hear the range of events or experiences that people will report.
  - *Conduct cognitive interviews.* Ask people a test question, then probe with follow-up questions about how they understood the question and what their answer meant.
  - *Audiotape test interviews during the pretest phase of a survey.* The researchers then review these audiotapes and systematically code them to identify problems in question wording or delivery. (pp. 104–129)

DISCUSSION QUESTIONS

1. What does poverty mean to you? Have you been using an absolute, relative, or objective standard when you think about poverty? Identify two examples of "poverty" and explain why they represent this concept. Compare your conceptualization with those of your classmates and what you find in a dictionary. Can you improve your conceptualization based on some feedback?

2. What questions would you ask to measure feelings of being "in" or "out" with regard to a group? Write five questions for an index and suggest response choices for each. How would you validate this measure using a construct validation approach? Can you think of a criterion validation procedure for your measure?

3. If you were given a questionnaire right now that asked you about your use of alcohol and illicit drugs in the past year, would you disclose the details fully? How do you think others would respond? What if the questionnaire was anonymous? What if there was a confidential ID number on the questionnaire so that the researcher could keep track of

who responded? What criterion validation procedure would you suggest for assessing measurement validity?

4. The questions in Figure 4.16 are selected from my survey of staff in a shelter for homeless persons (Schutt and Fennell 1992). First, identify the level of measurement for each question. Then, rewrite each question so that it measures the same variable but at a different level. For example, you might change the question that measures seniority at the ratio level (in years, months, and days) to one that measures seniority at the ordinal level (in categories). Or you might change a variable measured at the ordinal level, such as highest grade in school completed, to one measured at the ratio level. For the variables measured at the nominal level, try to identify at least two underlying quantitative dimensions of variation and write questions to measure variation along these dimensions. For example, you might change the question, "What is your current job title?" to two questions that ask about the pay in the respondent's current job and the extent to which the job is satisfying.

FIGURE 4.16 ■ Selected Shelter Staff Survey Questions

1. What is your current job title? _____

2. What is your current employment status?

   Paid, full-time _____ 1
   Paid, part-time (less than 30 hours per week) _____ 2

3. When did you start your current position? _____ / _____ / _____
   Month Day Year

4. In the past month, how often did you help guests deal with each of the following types of problems? (Circle one response on each line.)

	Very often						Never
Job training/placement	1	2	3	4	5	6	7
Lack of food or bed	1	2	3	4	5	6	7
Drinking problems	1	2	3	4	5	6	7

5. How likely is it that you will leave this shelter within the next year?

   Very likely _____ 1
   Moderately _____ 2
   Not very likely _____ 3
   Not likely at all _____ 4

6. What is the highest grade in school you have completed at this time?

   First through eighth grade _____ 1
   Some high school _____ 2
   High school diploma _____ 3
   Some college _____ 4
   College degree _____ 5
   Some graduate work _____ 6
   Graduate degree _____ 7

7. Are you a veteran?

   Yes _____ 1
   No _____ 2

Source: Based on Schutt 1992:7–10, 15, 16. Results reported in Schutt, R. K., and M. L. Fennell. 1992. "Shelter Staff Satisfaction With Services, the Service Network and Their Jobs." *Current Research on Occupations and Professions* 7:177–200.

What are the advantages and disadvantages of phrasing each question at one level of measurement rather than another? Do you see any limitations on the types of questions for which levels of measurement can be changed?

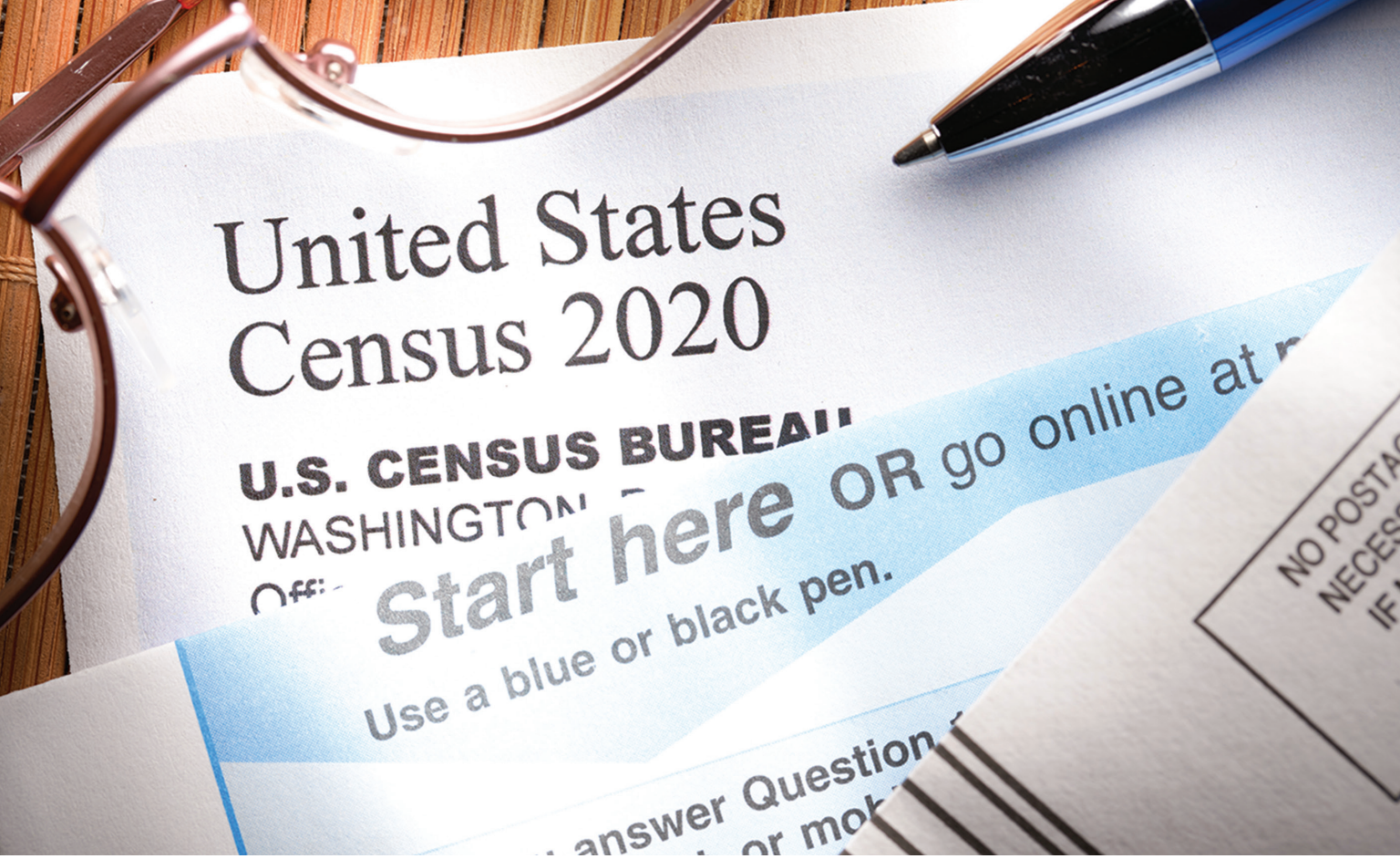
SPSS EXERCISES

1. View the variable information for the variables age_2, childs_2, partyid_2, socbar_2, race_2, and income16_2 in the GSS2020 file. Click on the "variable list" icon or choose Utilities/Variables from the menu. Choose partyid_2, then socbar_2. At which levels (nominal/categorical, ordinal, interval, ratio) are each of these variables measured? (By the way, DK means "Don't Know," NA means "No Answer," and NAP means "Not Applicable.")

2. Review the actual questions used to measure four of the variables in Question 1 or in your hypotheses in Chapter 2's SPSS exercises (Question 3). You can find most GSS questions at the Data Explorer website for the General Social Survey (gssdataexplorer.norc.org). Name the variable that you believe each question measures (search under "keyword"). Discuss the face validity and content validity of each question as a measure of its corresponding variable. Explain why you conclude that each measure is valid or not.

3. Fechild_2, fepresch_2, fefam_2, and meovrwrk_2 are part of an index involving the following question: Please read the following statements and indicate whether you strongly agree, agree, disagree, or strongly disagree with each statement:
   - **a.** A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.
   - **b.** A preschool child is likely to suffer if his or her mother works.
   - **c.** It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family.
   - **d.** Family life often suffers because men concentrate too much on their work.

   Now answer the following questions:
   - **i.** What is the concept being measured by this index?
   - **ii.** Do you agree that each of these variables belongs in the index? Explain.
   - **iii.** What additional variables would you like to see included in this index?



©iStockphoto.com/Bill Oxford

5 SAMPLING AND GENERALIZABILITY

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

5.1 Summarize the process of probability-based sampling

5.2 Describe the circumstances that make sampling unnecessary and the reasons they are rare

5.3 Explain why probability-based sampling is needed to obtain a representative sample

5.4 Define the different probability sampling methods

5.5 Defend the value of nonprobability sampling methods

5.6 Assess how sampling error is affected by sample size, the heterogeneity of the population, and the fraction of the population included in the sample

A common technique in journalism to increase interest in a story is to highlight the experiences of one or more individuals or communities that readers can relate to. For instance, *New York Times* reporter Michael Wines (2020) began a story on challenges the Census Bureau encountered in the 2020 Census by contrasting census completion rates in two communities in Michigan. In Royal Oak, a "comfortable" suburban community outside of Detroit, seven in ten households had completed their census form by July, with almost all doing so online. In Delray, a "faded industrial community" twenty miles away, half the households did not have Internet service and only two in ten had completed their census form. Wines then summarizes the multiple challenges encountered in a census that began just as the pandemic locked down the nation: initial delays in knocking on doors to encourage response in lower-income communities with historically lower participation rates; cancellation of summer festivals to publicize the civic benefits of the census; special efforts to overcome fears about participation due to the Trump administration's anti-immigrant policies; fewer phone calls due to social distancing requirements in bureau offices; slowdowns at the postal service.

The contrast in the early response to the census survey in Royal Oak and Delray highlights the type of difficulties faced in survey research as well as the special sampling challenges created by the pandemic. Since the results of the U.S. Census are used to draw political districts as well as to allocate funds for many public programs, the different response rates in these two communities remind us of the importance of designing methods to maximize generalizability of the findings to the entire population of interest. If poorer communities and communities of color are underrepresented in census results, people who need support programs the most will be less likely to receive them and have fewer representatives in Congress. However, without other information, we don't know whether the experience of these two communities with the census itself provides a "sample" that can be generalized to most communities in Michigan, most communities nationwide, or whether such experiences have occurred in other decades. In other words, we don't know just from the example how generalizable the stories from these two communities are, and if we don't have confidence in generalizability, then the validity of this comparison of participation in the 2020 Census is suspect. Because we don't know whether their situation was widely shared or unique, we cannot really judge what the contrast between Royal Oak and Delray tells us about the social world.

In this chapter, you will learn about sampling methods, the procedures that primarily determine the generalizability of research findings. I first review the rationale for using sampling in social research and consider two circumstances when sampling may not seem necessary, including when a complete census of the population is conducted. I then turn to specific sampling methods and when they are most appropriate. That discussion is followed by a discussion of sampling distributions, which introduces you to the logic of statistical inference—that is, how to determine the likelihood that features of a population can be inferred from a sample of that population. By the chapter's end, you should understand what you need to know to evaluate the generalizability of a study as well as what choices you need to make when designing a sampling strategy.

SAMPLE PLANNING

You have encountered the problem of generalizability in each of the studies you have read about in previous chapters. For example, Chad Ke (Chapter 1) studied student reflection papers in one graduate class; could his findings be generalized to all graduate students or even just those in higher education leadership courses like his? The U.S. Public Health Service researchers who began the Tuskegee Syphilis Study in 1932 (Chapter 3) justified it in part with the unfounded belief that prior research on the course of syphilis in white people could not be generalized to Black people. How confident can we be in generalizing the finding about level of support for defunding the police from the 2020 Associated Press/NORC poll of a sample of Americans to the entire adult population or from the Fox News poll of a sample of voters in the same year (Chapter 4) to all voters? Whether you are designing a sampling strategy or evaluating someone else's findings, you should understand how and why researchers decide to sample and what the consequences of these decisions are for the generalizability of the study's findings.

The Purpose of Sampling

The purpose of sampling is to generate a set of individuals or other entities that give you a valid picture of all such individuals or other entities. That is, a sample is a subset of the larger set of individuals or other entities in which you are interested. If you have done a good job of sampling, you will be able to generalize what you have learned from the subset to the larger set from which it was selected.

As researchers, we call the set of individuals or other entities to which we want to generalize our findings the **population**. For example, a city government may want to describe the city's entire adult homeless population. If, as is usually the case, the government does not have the time or resources to survey all homeless individuals in the city, it may fund a survey of a subset of these individuals. This subset of the population of interest is a **sample**. The individual members of this sample are called **elements**, or elementary units.

Define Sample Components and the Population

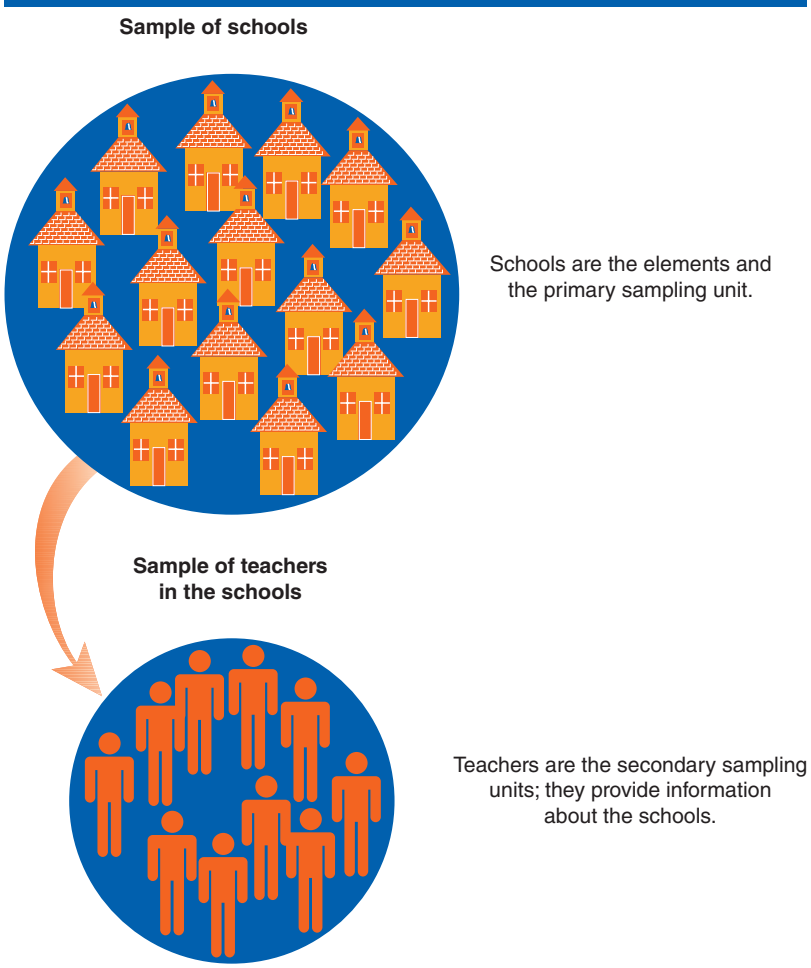
In many studies, we sample directly from the elements in the population of interest. We may survey a sample of the entire population of students in a school, based on a list obtained from the registrar's office. This list, from which the elements of the population are selected, is termed the **sampling frame**. The students who are selected and interviewed from that list are the elements. We have to review the sampling frame carefully to determine whether some elements in the population are missing (such as new students who have not yet completed some forms), whether some elements are included that are not part of the target population (such as students who have

moved away but have not yet been dropped from the registrar's list), and whether any elements are duplicated in the list (Kalton 2020:5).

In some studies, the entities that can be reached easily are not the same as the elements from which we want information, but they include those elements. For example, you may have a list of households but not a list of all the persons in a town, when the adults are the elements that you want to sample. In this situation, you could draw a sample of households and then identify the adult individuals in these households. The households are termed **enumeration units**, and the adults in the households are the elements (Levy and Lemeshow 1999:13–14).

Sometimes, the individuals or other entities from which we collect information are not actually the elements in our study. For example, a researcher might sample schools for a survey about educational practices and then interview a sample of teachers in each sampled school to obtain the information about educational practices. Both the schools and the teachers are termed **sampling units** because the researcher samples from both (Levy and Lemeshow 1999:22). The schools are selected in the first stage of the sample, so they are the *primary sampling units* (in this case, they are also the elements in the study). The teachers are *secondary sampling units* (but they are not elements because they are used to provide information about the entire school) (see Figure 5.1).

FIGURE 5.1 ■ Sample Components in a Two-Stage Study


Sample of schools
Schools are the elements and the primary sampling unit.
Sample of teachers in the schools
Teachers are the secondary sampling units; they provide information about the schools.

Source: Based on information from Levy and Lemeshow (1999).

It is important to know exactly what population a sample can represent when you select or evaluate sample components. In a survey of "adult Americans," the general population may reasonably be construed as all residents of the United States who are at least 21 years old. But always be alert to ways in which the population may have been narrowed by the sample selection procedures. For example, if only English-speaking residents were surveyed, then the population we studied is English-speaking residents of the United States, not "all residents" of the United States. The population for a study is the aggregation of elements that we sample from, not some larger aggregation that we really wish we could have studied. The ideal population to which we seek to generalize can be termed the **target population**—a set of elements to which the researcher would ideally like to generalize any study findings but that may differ from the actual population studied due to practical or other constraints (Kalton 2020:2).

Some populations, such as homeless persons, are not identified by a simple criterion, such as a geographic boundary or an organizational membership. Clear definition of such a population is difficult but quite necessary. Anyone should be able to determine just what population was studied. For example, several studies of homeless persons in urban areas defined the population of interest as, "people who had no home or permanent place to stay of their own (meaning they rented or owned it themselves) and no regular arrangement to stay at someone else's place" (Burt 1996:18). But even this explicit definition leaves some questions unanswered: What is a "regular arrangement"? How permanent does a "permanent place" have to be? In a study of homeless persons in Chicago, Sosin et al. (1988) answered these questions in their definition of the population of interest:

> We define the homeless as: those current[ly] residing for at least one day but for less than fourteen with a friend or relative, not paying rent, and not sure that the length of stay will surpass fourteen days; those currently residing in a shelter, whether overnight or transitional; those currently without normal, acceptable shelter arrangements and thus sleeping on the street, in doorways, in abandoned buildings, in cars, in subway or bus stations, in alleys, and so forth; those residing in a treatment center for the indigent who have lived at the facility for less than 90 days and who claim that they have no place to go, when released. (p. 22)

This definition reflects accurately Sosin et al.'s concept of homelessness and allows researchers in other locations or at other times to develop procedures for studying a comparable population.

In a study of unauthorized and unsuccessful Mexican migrants' experiences with border crossing, Martinez et al. (2017:253) defined the population for their study at a private shelter in Mexico as Spanish-speaking individuals 18 years or older who had

> attempted an undocumented crossing along the Sonora–Arizona border, been apprehended by a US authority, and repatriated to Mexico, all within the past six months.

The more complete the definition of the population from which a sample was selected, the more precise generalizations can be.

SAMPLE GENERALIZABILITY

Once we have defined the population from which we will sample, we need to determine the scope of the generalizations we will make from our sample. Do you recall from Chapter 2 the two different meanings of generalizability?

1. *Can the findings from a sample of the population be generalized to the population from which the sample was selected?*

   Can findings from the 2020 Associated Press/NORC poll of a sample of Americans about defunding the police be generalized to the entire adult population? This type of generalizability was defined as *sample generalizability* inChapter 2.

2. *Can the findings from a study of one population be generalized to another, somewhat different population?*

   Could the effect of a police officer beating in Milwaukee that Matthew Desmond, Andrew Papachristos, and David Kirk (2016) (Chapter 2) found on subsequent calls to 911 be generalized to Black communities in other cities? Do findings from a laboratory study about obedience to authority at an elite northeastern U.S. college in the 1960s differ from those that would be obtained today at a commuter college in the Midwest (Milgram 1963)? What is the generalizability of the results from a survey of repatriated migrant border crossers in one shelter in Nogales, Sonora, to the entire Tucson Sector (most of Arizona) of the U.S. Border Patrolor to migrants who were returned to other shelters (Martinez et al. 2017)? This type of generalizability was defined as *cross-population generalizability* inChapter 2.

This chapter focuses attention primarily on the problem of sample generalizability: Can findings from a sample be generalized to the population from which the sample was drawn? This is really the most basic question to ask about a sample, and social research methods provide many tools with which to address it.

Sample generalizability depends on sample quality, which is determined by the amount of **sampling error**—the difference between the characteristics of a sample and the characteristics of the population from which it was selected. The larger the sampling error, the less representative the sample—and thus the less generalizable the findings. To assess sample quality when you are planning or evaluating a study, ask yourself these questions:

- From what population were the cases selected?
- What method was used to select cases from this population?
- Do the cases that were studied represent, in the aggregate, the population from which they were selected?

Cross-population generalizability (also called "external validity") involves quite different considerations. Researchers are engaging in cross-population generalizability when they project their findings onto different groups or target populations than those they have studied. If we seek to generalize our findings from a sample of residents in one city to all cities in the United States, there is a large gap between the population we studied and the target population. We must carefully consider the validity of claims that findings can be generalized from the population studied to other groups, geographic areas, cultures, or times.

The validity of cross-population generalizations cannot be tested empirically, except by conducting more research in other settings, but it is often a concern in qualitative studies and experimental research that does not attempt to sample from a defined population. I discuss the problem of cross-population generalizability later in this chapter, in the context of nonprobability sampling methods that are often used in qualitative research, and then again in Chapter 7, which addresses experimental research, in Chapter 10, which addresses qualitative methods, and in Chapter 15, which discusses methods for studying different societies.

Assess the Diversity of the Population

Sampling is unnecessary if all the units in the population are identical. Physicists don't need to select a representative sample of atomic particles to learn about basic physical processes. They can study a single atomic particle because it is identical to every other particle of its type. Similarly, biologists don't need to sample a particular type of plant to determine whether a given chemical has toxic effects on that particular type. The idea is "If you've seen one, you've seen 'em all."

What about people? Certainly, all people are not identical (nor are other animals, in many respects). Nonetheless, if we are studying physical or psychological processes that are the same among all people, sampling is not needed to achieve generalizable findings. Psychologists and social psychologists often conduct experiments on college students to learn about processes they think are identical across individuals. They believe that most people would have the same reactions as the college students if they experienced the same experimental conditions. Field researchers who observe group processes in a small community sometimes make the same assumption.

Unfortunately, there's no way to know for sure if the processes being studied are identical across all people, places, and times. In fact, experiments can give different results depending on the type of people who are studied or the conditions for the experiment. Milgram's (1965) classic experiments on obedience to authority, which you studied in Chapter 3, illustrate this point very well. Stanley Milgram concluded from his initial experiments with male volunteers in New Haven, Connecticut, that people are very obedient to authority. Similar results were obtained in many replications of the Milgram experiments—that is, when the experimental conditions and subjects were similar to those Milgram had studied. But other studies showed that given somewhat different conditions, such as another "subject" in the room who refused to administer the shocks, subjects were likely to resist authority (Miller 1986).

So what do the initial experimental results tell us about how people will react to an authoritarian movement in the real world, when conditions are not so carefully controlled? Generalizing the results of single experiments is always risky because such research often studies a small number of people who are not selected to represent any particular population.

But what if your goal is not to learn about individuals but about the culture or subculture in a society or group? The logic of sampling does not apply if the goal is to learn about culture that is shared across individuals:

> When people all provide the same information, it is redundant to ask a question over and over. Only enough people need to be surveyed to eliminate the possibility of errors and to allow for those who might diverge from the norm. (Heise 2010:15)

If you are trying to describe the culture of a group or society, you may choose individuals for the survey based on their knowledge of the culture, not as representatives of a population of individuals (Heise 2010:16). In this situation, what is important about the individuals surveyed is what they have in common, not their diversity.

Keep these exceptions in mind, but the main point is that social scientists rarely can skirt the problem of demonstrating the generalizability of their findings. If a small sample has been studied in an experiment or a field research project, the study should be replicated in different settings. The social world and the people in it are just too diverse to be considered *identical units* in most respects. Social psychological experiments and small field studies have produced good social science, but they need to be replicated in other settings, with other subjects, to claim any generalizability.

Consider a Census

In some circumstances, it may be feasible to skirt the issue of generalizability by conducting a **census**—studying the entire population of interest—rather than drawing a sample. Article 1 of the United States Constitution requires that the federal government conduct a census every 10 years and use the results to apportion federal taxes. A census can also be used in a study of all the employees (or students) in small organizations, studies of the United States comparing all 50 states (so states are the elements), and studies of the entire population of a particular type of organization. However, in comparison with the U.S. Census and similar efforts in other nations, states, and cities, the population that is studied in these other censuses is relatively small.

Social scientists don't often attempt to collect data from all the members of some large population because of the time and expense it requires—and because selecting a sample for study often yields data that is just as useful. To conduct the 2020 Census for the United States population of about 330 million in more than 140 million housing units (https://www.census.gov/programs-surveys/decennial-census/decade/2020/about.html), the U.S. Census Bureau spent about $15.6 billion to ask just nine questions (seven of them repeated with reference to each of six household members) (Decennial Census | U.S. GAO). The bureau partnered with more than 400,000 other organizations, offered the census questionnaire in 13 different languages with online, phone, and in-person options, and checked the responses against a file of addresses throughout the United States to ensure complete coverage and elimination of duplicate responses (U.S. Census Bureau 2021). All persons residing in the United States on census day are supposed to be counted, including citizens and permanent residents, documented and undocumented immigrants and homeless persons, and people living in group settings like military bases, prisons, college dorms, and hospitals (Donnelly 2019:3). Since the resulting data are released in public reports and made available for analysis by other researchers (without information identifying individuals), the census provides a lot of value for the taxes that fund it, but this is simply not a feasible approach for focused social science research projects about issues not covered in the official census.

Even if the population of interest for a survey is a small town of 20,000 or students in a university of 10,000, researchers will have to sample. The costs of surveying "just" thousands of individuals exceed by far the budgets for most research projects. In fact, not even the U.S. Census Bureau can afford to include many important questions related to ongoing government concerns in the decennial census. So it draws a sample. Every year, a sample of 3.5 million households is sent a long-form (with about 50 questions in 2020 for the primary household respondent) (https://www.census.gov/programs-surveys/acs.html). This more detailed sample survey began in 2005 as the American Community Survey and replaces what formerly was a long-form of the census that was administered to one sixth of the population at the same time as the regular census (Donnelly 2019; U.S. Census Bureau 2010d). Other survey programs of the U.S. Census Bureau are described in Chapter 14.

The fact that it is hard to get people to complete a survey is another reason why conducting a census can be costly. Even the U.S. Census Bureau (1999) must make multiple efforts to increase the rate of response, despite the federal law requiring all citizens to complete their census questionnaire. After up to seven efforts to encourage response, two thirds of households responded to the 2020 Census online, by mail, or by phone, while another third responded in an in-person interview or had their characteristics estimated by a neighbor or through checking other government records (https://www.census.gov/library/visualizations/interactive/2020-census-total-response-rates-by-state.html).

Even after all that, we know from the 2000 U.S. Census that some groups are still likely to be underrepresented (Armas 2002; Holmes 2001a), including minority groups (Kershaw 2000), impoverished cities (Zielbauer 2000), well-to-do individuals in gated communities and luxury buildings (Langford 2000), and even college students (Abel 2000). The number of persons missed in the 2000 Census was estimated to be between 3.2 and 6.4 million (U.S. Census Bureau 2001). The effort by the Trump administration to add a question on citizenship to the 2020 Census seemed to many census experts likely to instill fear of participating among undocumented immigrants and within their larger communities. However, the Supreme Court prohibited the addition of the citizenship question in June 2019, (Donnelly 2019:11). An expert panel commissioned by the American Statistical Association concluded, in an early analysis, that the 2020 Census, conducted during the pandemic and after the controversy over the citizenship question, achieved as accurate a count as the highly regarded 2010 Census (Wines 2021), although it continued to undercount Black, Hispanic, and Native American residents--missing between 3 and 6 out of every 100 (Wines and Cramer 2022).

The average survey project has far less legal and financial backing, and thus an adequate census is not likely to be possible. Consider the problems of conducting a census in Afghanistan. The first census in 23 years was conducted by the country's Central Statistics Office in 2003 and 2004, interrupted by snow that cut off many districts for 6 or 7 months. Teams of census takers carried tents, sleeping bags, and satellite phones as they trekked into remote mountainous provinces. An accompanying cartographer identified the location of each village using global positioning systems (GPS) (Gall 2003:A4). Many other countries have replaced their census with annual sample surveys, estimates from administrative data, or population registers of some kind (Donnelly 2019:9). In smaller groups or organizations, however, a complete census may be a viable strategy.

In most survey situations, it is much better to survey only a limited number from the total population so there are more resources for follow-up procedures that can overcome reluctance or indifference about participation. (I give more attention to the problem of nonresponse in Chapter 8.)

PROBABILITY AND REPRESENTATIVENESS

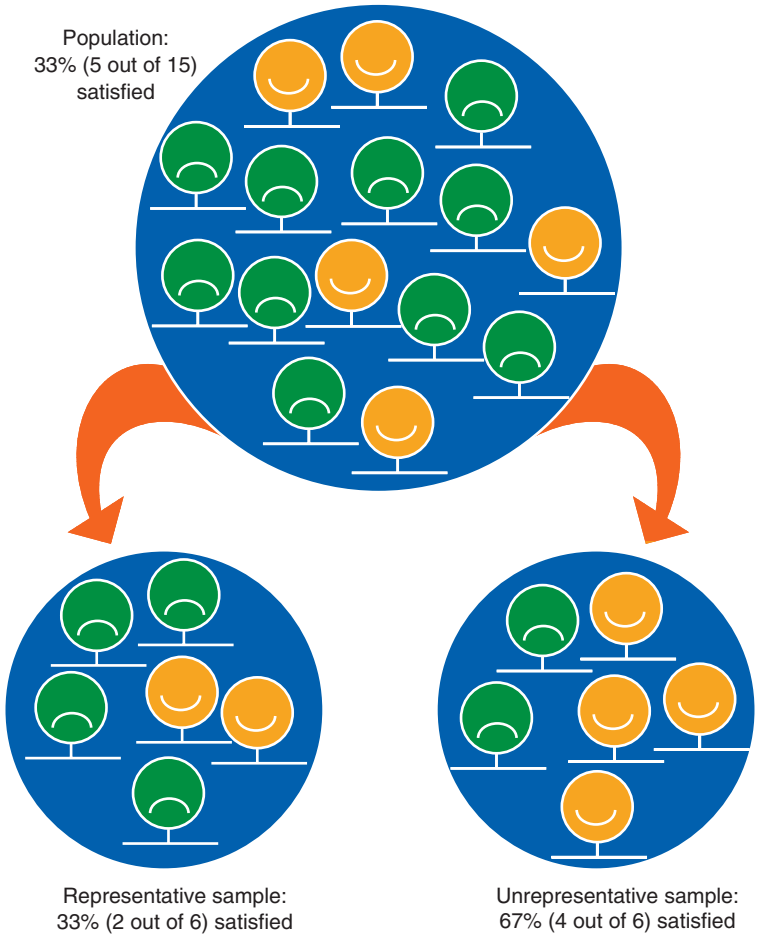
To generalize with some confidence from a sample to a population, a **representative sample** must be selected from the population to which generalizations are sought (see Figure 5.2).

We can now study more systematically the features of samples that make them more or less likely to represent the population from which they are selected. The most important distinction that needs to be made about the samples is whether they are based on a probability or a nonprobability sampling method (Kalton 2020:2). Sampling methods that allow us to know in advance how likely it is that any element of a population will be selected for the sample are termed **probability sampling methods**. Sampling methods that do not let us know in advance the likelihood of selecting each element are termed **nonprobability sampling methods**.

Probability sampling methods rely on a random, or chance, selection procedure, which is, in principle, the same as flipping a coin to decide which of two people "wins" and which one "loses." Heads and tails are equally likely to turn up in a coin toss, so both persons have an equal chance of winning. That chance, their **probability of selection**, is 1 out of 2, or .5.

Flipping a coin is a fair way to select one of two people because the selection process harbors no systematic bias. You might win or lose the coin toss, but you know that the outcome was due

FIGURE 5.2 ■ Representative and Unrepresentative Samples



simply to chance, not to bias. For the same reason, a roll of a six-sided die is a fair way to choose one of six possible outcomes (the odds of selection of any one outcome are 1 out of 6, or .17). Dealing out a hand after shuffling a deck of cards is a fair way to allocate sets of cards in a poker game (the odds of each person getting a particular outcome, such as a full house or a flush, are the same). Similarly, state lotteries use a random process to select winning numbers. Thus, the odds of winning a lottery, the probability of selection, are known, even though they are very much smaller (perhaps 1 out of 1 million) than the odds of winning a coin toss.

There is a natural tendency to confuse the concept of **random sampling**, in which cases are selected only based on chance, with a haphazard method of sampling. On first impression, "leaving things up to chance" seems to imply not exerting any control over the sampling method. But to ensure that nothing but chance influences the selection of cases, the researcher must proceed methodically, leaving nothing to chance except the selection of the cases themselves. The researcher must follow carefully controlled procedures if a purely random process is to occur. When reading about sampling methods, do not assume that a random sample was obtained just because the researcher used a random selection method at some point in the sampling process. Look for two problems in particular: selecting elements from an incomplete list (sampling frame) of the total population and failing to obtain an adequate response rate.

If the sampling frame is incomplete, a sample selected randomly from that list will not really be a random sample of the population. You should always consider the adequacy of the sampling frame. Even for a simple population such as a university's student body, the registrar's list is likely to be at least a bit out-of-date at any given time. For example, some students will have dropped out, but their status will not yet be officially recorded. Although you may judge the amount of error introduced in this situation to be negligible, the problems are greatly compounded for a larger population. The sampling frame for a city, state, or nation is always likely to be incomplete because of moves into and out of the area. Even unavoidable omissions from the sampling frame can result in some sample bias.

An inclusive sampling frame may still yield systematic bias if many sample members cannot be contacted or refuse to participate. Nonresponse is a major hazard in survey research because **nonrespondents** are likely to differ systematically from those who take the time to participate. You should not assume that findings from a randomly selected sample will be generalizable to the population from which the sample was selected if the rate of nonresponse is considerable (much above 30%). Unfortunately, response rates have been declining in recent decades (for reasons discussed in Chapter 8). The response rates in major national surveys conducted with in-person interviews have dropped from over 90% in the 1990s to about 70% in the 2010s (Kalton 2020:88). Response rates are now often much lower in mail and phone surveys.

Probability sampling methods are those in which the probability of selection is known and is not zero (so there is some chance of selecting each element). These methods randomly select elements and therefore have no **systematic bias**; nothing but chance determines which elements are included in the sample. This feature of probability samples makes them much more desirable than nonprobability samples when the goal is to generalize to a larger population.

Although a random sample has no systematic bias, it will certainly have some sampling error resulting from chance. The *probability* of selecting a head is .5 in a single toss of a coin and in 20, 30, or however many tosses of a coin you like. But it is perfectly possible to toss a coin twice and get a head both times. The random "sample" of the two sides of the coin is selected in an unbiased fashion, but it still is unrepresentative. Imagine selecting randomly a sample of 10 people from a population comprising 50 men and 50 women. Just by chance, can't you imagine that these 10 people selected based on chance include seven women and only three men? Fortunately, we can determine mathematically the likely degree of sampling error in an estimate based on a random sample (as we'll discuss later in this chapter)—assuming the sample's randomness has not been destroyed by a high rate of nonresponse or by poor control over the selection process.

In general, both the size of the sample and the homogeneity (sameness) of the population affect the degree of error because of chance; the proportion of the population that the sample represents does not have much effect on the degree of sampling error. To elaborate,

- *The larger the sample, the more confidence we can have in the sample's representativeness.* If we randomly pick 5 people to represent the entire population of our city, our sample is unlikely to be very representative of the entire population in age, gender, race, attitudes, and so on. But if we randomly pick 100 people, the odds of having a representative sample are much better; with a random sample of 1,000, the odds become very good indeed.
- *The more homogeneous the population, the more confidence we can have in the representativeness of a sample of a particular size.* Let's say we plan to draw samples of 50 from each of two communities to estimate mean family income. One community is quite diverse, with family incomes varying from $12,000 to $85,000. In the other, more

homogeneous community, family incomes are concentrated in a narrow range, from $41,000 to $64,000. The estimated mean family income based on the sample from the more homogeneous community is more likely to be representative than is the estimate based on the sample from the more heterogeneous community. With less variation to represent, fewer cases are needed to represent the homogeneous community.

- *The fraction of the total population that a sample contains does not affect the sample's representativeness unless that fraction is large.* We can regard any sampling fraction less than 2% with about the same degree of confidence (Sudman 1976:184). In fact, sample representativeness is not likely to increase much until the sampling fraction is quite a bit higher. Other things being equal, a sample of 1,000 from a population of 1 million (with a sampling fraction of 0.001, or 0.1%) is much better than a sample of 100 from a population of 10,000 (although the sampling fraction for this smaller sample is 0.01, or 1%, which is 10 times higher). Larger size makes sample representativeness more likely, not a higher proportion that the sample represents of the whole (except as just noted).

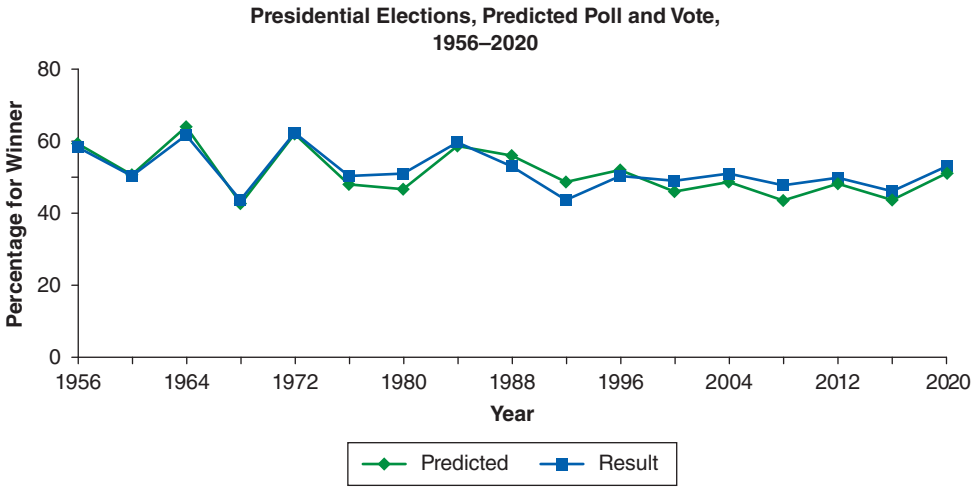
Polls to predict presidential election outcomes illustrate both the value of random sampling and the problems that it cannot overcome. Prior to the use of random sampling in polling, there were some major snafus. In 1936, a *Literary Digest* poll predicted that Republican Alfred M. Landon would defeat Democrat Franklin Delano Roosevelt, the incumbent President, in a landslide, but instead, Roosevelt took 63% of the popular vote. The problem? The *Digest* mailed out 10 million mock ballots to people listed in telephone directories, automobile registration records, voter lists, and so on. But in 1936, during the Great Depression, only relatively wealthy people had phones and cars, and they were more likely to be Republican. Furthermore, only 2,376,523 completed ballots were returned, and a response rate of only 24% leaves much room for error. Of course, this poll was not designed as a random sample, so the appearance of systematic bias is not surprising. Gallup predicted the 1936 election results accurately with a more systematically selected sample of just 3,000 that avoided much of the bias in the Literary Digest poll (although they did not yet use random sampling) (Bainbridge: 1989:43–44).

In 1948, pollsters mistakenly predicted that Republican Thomas E. Dewey would beat incumbent Democrat President Harry S. Truman, based on the sampling method that George Gallup had used successfully since 1934. The problem was that pollsters stopped collecting data several weeks before the election, and in those weeks, many people changed their minds (Kenney 1987). The sample was systematically biased by underrepresenting shifts in voter sentiment just before the election. This experience convinced Gallup to use only random sampling methods (as well as to continue polling until the election).

In most subsequent presidential elections, pollsters have predicted accurately the outcomes of the actual votes by using random sampling and phone interviewing to learn for which candidate the likely voters intend to vote. Figure 5.3 shows how close these sample-based predictions have been in the past 17 contests. The exceptions were the 1980 and 1992 elections, when third-party candidates had an unpredicted effect. Otherwise, the small discrepancies between the votes predicted through random sampling and the actual votes can be mostly attributed to random error.

The Gallup poll did not do as well in predicting the results of the 2008 presidential election. The final 2008 Gallup prediction was that Barack Obama would win with 55% to John McCain's 44% (Gallup 2011). The race turned out a bit closer, with Obama winning by 53% to McCain's 46%, with other polling organizations closer to the final mark (the Rasmussen and Pew polls were exactly on target) (Panagopoulos 2008). However, the overall rate of accuracy has been impressive.

FIGURE 5.3 ■ Presidential Election Outcomes: Predicted and Actual



Sources: Peters, Gerhard. n.d. "Election Year Presidential Preference Polling and Voting Outcomes." The American Presidency Project. John T. Woolley and Gerhard Peters (eds.). Santa Barbara, CA: University of California. Downloaded April, 13, 2022: https://www.presidency.ucsb.edu/node/323909.

Wikipedia. 2022. Historical Polling for United States Presidential Elections. Online. Downloaded April 13, 2022. https://en.wikipedia.org/wiki/Historica_polling_for_United_States_presidential_elections?msclkid=e2932996b5611eca8075e3f04f64e7a.

But what about the 2016 and 2020 presidential elections, with Donald Trump's surprising win in 2016 and surprisingly close finish in 2020 (relative to expectations based on most polls)? Although Gallup did not conduct presidential polls in 2016, the composite prediction from 13 other national polls in the week before that election was that Hillary Clinton would win the popular vote by 3.1 points—a gap of only 1 point from her advantage of 2.1% in the national vote total calculated in November (Clement and Craighill 2015; Newport 2016). What polling got wrong were predictions in the so-called battleground states in which undecided voters broke for Trump in the last days and Trump supporters turned out at higher than predicted rates, while pollsters didn't adequately take account of educational differences in political preferences when they adjusted for likely turnout (college-educated voters favored Clinton by a 25-point margin) (Cohn 2017). The relatively accurate prediction from national polls of Clinton's popular margin, therefore, did not accurately forecast the large advantage for Trump in the electoral college due to these battleground states.

Although Joe Biden's win in 2020 was decisive in the Electoral College (306 to 232), his margin in the popular vote of 51.3% to 46.9% was much closer than the 8% Biden advantage predicted by presidential polls (although that gap narrowed in the days before the election) (https://www.270towin.com/2020-polls-biden-trump/national/). The additional details in this chapter's Research in the News story emphasize just how important nonresponse bias can be for explaining sample bias like this.

RESEARCH IN THE NEWS
"A BLACK EYE": WHY POLITICAL POLLING MISSED THE MARK. AGAIN.

Most polls conducted before the 2020 presidential election correctly predicted that Joe Biden would defeat then-president Donald Trump, but by eight percentage points rather than Biden's actual winning margin in the national vote total of four percentage points. A four-point

margin of error may not sound too bad in predicting the popular vote, but the details were much worse (although Biden's margin was much higher in the electoral college). Most polls predicted a landslide win by Biden in many states that he won by less than two percentage points. Both Democratic- and Republican-affiliated polling firms underestimated support for Republicans. A key problem is that political polls have not been able to eliminate bias in their samples. Few people respond to phone polls in recent years (average phone polling response is now only 6%, thanks in part to caller ID, and those who distrust government more are less likely to respond to polls—and nowadays distrust of government is higher among Republicans. Political polls faced other problems in the 2020 presidential race, but sample bias seems to have been the most important.

For Further Thought?

1. Some polling firms plan to include samples of text messages, Facebook messages, and other forms of nonverbal communication in their election outcome predictions. What do you see as the advantages and disadvantages of this approach?
2. J. Ann Selzer, a pollster in Iowa, has found that keeping polls very short increases responses of people who are less willing to respond. What strategies do you think would increase the response rate to political polls?

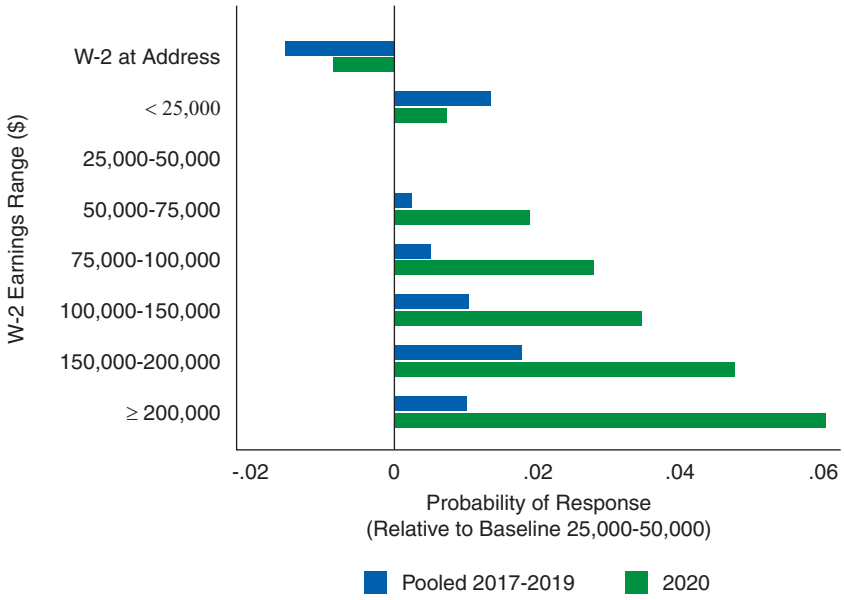
News source: Leonhardt, David. 2020. "'A Black Eye': Why Political Polling Missed the Mark. Again." *The New York Times,* November 12. Updated Nov. 30, 2020. https://www.nytimes.com/2020/11/12/us/politics/election-polls-trump-biden.html.

Because they do not disproportionately exclude or include certain groups within the population, truly random samples avoid systematic bias in the selection process. However, when some types of people are more likely to refuse to participate in surveys or are less likely to be available for interviews or are less likely to disclose their sentiments, systematic bias can still creep into the sampling process (Brenner 2021). The coronavirus pandemic resulted in a dramatic increase in this type of nonresponse bias. Even in the Current Population Survey, which respondents are legally required to complete, nonresponse increased much more among those with lower incomes compared to those with higher incomes (see Figure 5.4). As a result, even surveys using random selection methods in 2020 (and likely also 2021) would have resulted in underrepresenting persons with lower income, as well as associated characteristics like education and citizenship.

In addition, random error will still influence the specific results obtained from any random sample, and of course, opinions that are in flux create uncertainty about changes between the time of data collection and the date of releasing findings.

As a result of these potential sources of error, even surveys designed to be representative of a population must be evaluated in relation to the details of their methodology and the sample achieved. For example, the Latino National Health and Immigration Survey designed by the Robert Wood Johnson Foundation Center for Health Policy at the University of New Mexico was conducted with a random sample of both landline and mobile phones selected from the 44 states (and Puerto Rico) with the highest number of Latino residents (in the United States). However, the response rate to the telephone sample was only 18%. In addition, the phone sample was supplemented with responses from a sample randomly selected from an ongoing national internet panel. All surveys were available in either English or Spanish (Sanchez Pedraza and Vargas 2015; Vargas et al. 2019).

FIGURE 5.4 ■ Probability of Response by Total W-2 Earnings at Address



Source: Rothbaum, Jonathan, and Adam Bee. 2021. "Coronavirus Infects Surveys, Too: Survey Nonresponse Bias and the Coronavirus Pandemic." U.S. Census Bureau Working Paper # SEHSD WP2020-10, p. 44.

Let's review sample quality with the questions posed earlier:

- From what population were the cases selected? The population surveyed is described generally as "Latino adults in the United States," although the phone sample was only from the 44 states (and Puerto Rico). This introduces some ambiguity into the definition of the actual population to which results can be generalized.
- What method was used to select cases from this population? Those interviewed by phone were selected randomly. Those invited to participate in the web version of the survey were selected randomly from a preexisting national Internet panel of Latino adults. The method of recruitment for the nternet panel is not explained in the report and publication cited.
- Do the cases that were studied represent, in the aggregate, the population from which they were selected? It is not possible to determine the confidence that can be placed in the representativeness of the sample from the information provided in the sources cited. The most important issues of concern are the low response rate in the phone survey and the lack of information about the selection of the participants in the internet panel.

PROBABILITY SAMPLING METHODS

The likely amount of random error will also vary with the specific type of random sampling method used. The four most common methods for drawing random samples are (1) simple random sampling, (2) systematic random sampling, (3) stratified random sampling, and (4) cluster sampling.

Simple Random Sampling

Simple random sampling requires some procedure that generates numbers or otherwise identifies cases strictly based on chance. As you know, flipping a coin or rolling a die can be used to identify cases strictly based on chance, but these procedures are not very efficient tools for drawing a sample. A **random number table**, such as the one in Appendix B, simplifies the process considerably. The researcher numbers all the elements in the sampling frame and then uses a systematic procedure for picking corresponding numbers from the random number table (see details in Appendix B). Alternatively, a researcher may use a lottery procedure in which each case number is written on a small card or otherwise marked on small identical objects like tiny balls, and then the cards or other objects are mixed up and the sample is selected from the cards or other objects.

When a large sample must be generated, these procedures are cumbersome. Fortunately, a computer program can easily generate a random sample of any size (actually, most computer programs use a process that generates what is called a "pseudorandom" sequence of numbers, but that difference has no practical effect in social science research). The researcher must first number all the elements to be sampled (the sampling frame) and then run the computer program to generate a random selection of the numbers within the desired range. The elements represented by these numbers are the sample.

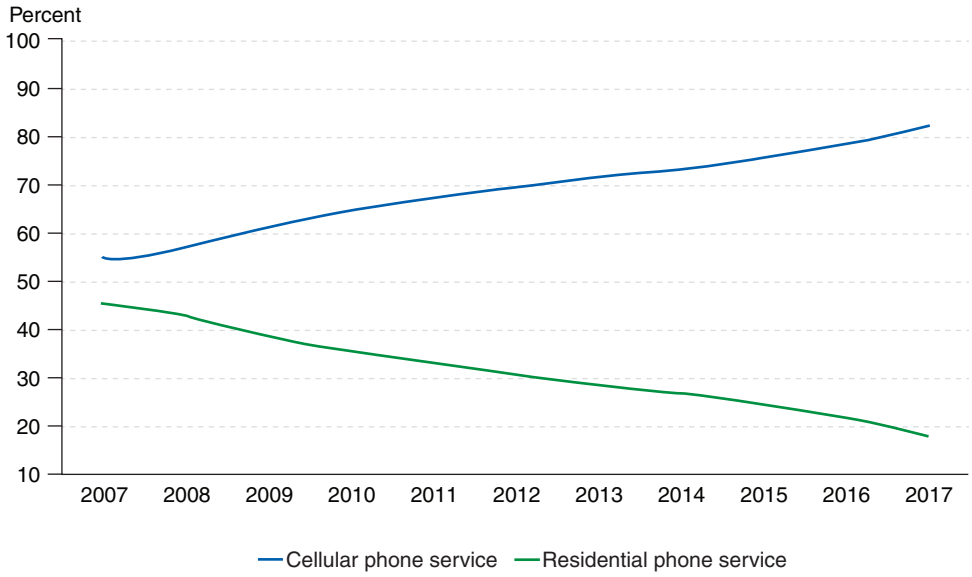
Organizations that conduct phone surveys often draw random samples using another automated procedure, called **random digit dialing**. A machine dials random numbers within the phone prefixes corresponding to the area in which the survey is to be conducted. Random digit dialing is particularly useful when a sampling frame is not available. The researcher simply replaces any inappropriate number (e.g., those that are no longer in service or that are for businesses) with the next randomly generated phone number.

As the fraction of the population that has only cell phones has increased (59% in 2019—see Figure 5.5), it has become essential to explicitly sample cell phone numbers as well as landline phone numbers (McGeeney and Keeter 2014; Woodie 2020). Compared with those who have a landline phone, those who use cell phones only tend to be younger; are more likely to be male, single, and Black or Hispanic; and are less likely to vote. As a result, failing to include cell phone numbers in a phone survey can introduce bias (Christian et al. 2010). For example, in a 2008 presidential election survey, those who used only cell phones were less likely to be registered voters than were landline users but were considerably more favorable to Obama than landline users (Keeter 2008). A survey just using landline users would have under-represented support for Obama.

The probability of selection in a true simple random sample is equal for each element. If a sample of 500 is selected from a population of 17,000 (i.e., a sampling frame of 17,000), then the probability of selection for each element is 500 to 17,000, or .03. Every element has an equal chance of being selected, just like the odds in a toss of a coin (1 to 2) or a roll of a die (1 to 6). Thus, simple random sampling is an *equal probability of selection method*, or EPSEM.

Simple random sampling can be done either with or without replacement sampling. In **replacement sampling**, each element is returned to the sampling frame after it is selected so that it may be sampled again. In sampling without replacement, each element selected for the sample is then excluded from the sampling frame (Kalton 2020:11). In practice, it makes no difference whether sampled elements are replaced after selection if the population is large, and the sample is to contain only a small fraction of the population. Random sampling with replacement is, in fact, rarely used.

FIGURE 5.5 ■ Percentage Allocated to Telephone Expenditures, by Cellular Photo Service and Residential Phone Service Expenditures, All Consumer Units, 2007–2017



Source: Creech, Brett, "Are most Americans cutting the cord on landlines?" *Beyond the Numbers: Prices & Spending* 8(7) (U.S. Bureau of Labor Statistics, May 2019), https://www.bls.gov/opub/btn/volume-8/are-most-americans-cutting-the-cord-on-landlines.htm.

Systematic Random Sampling

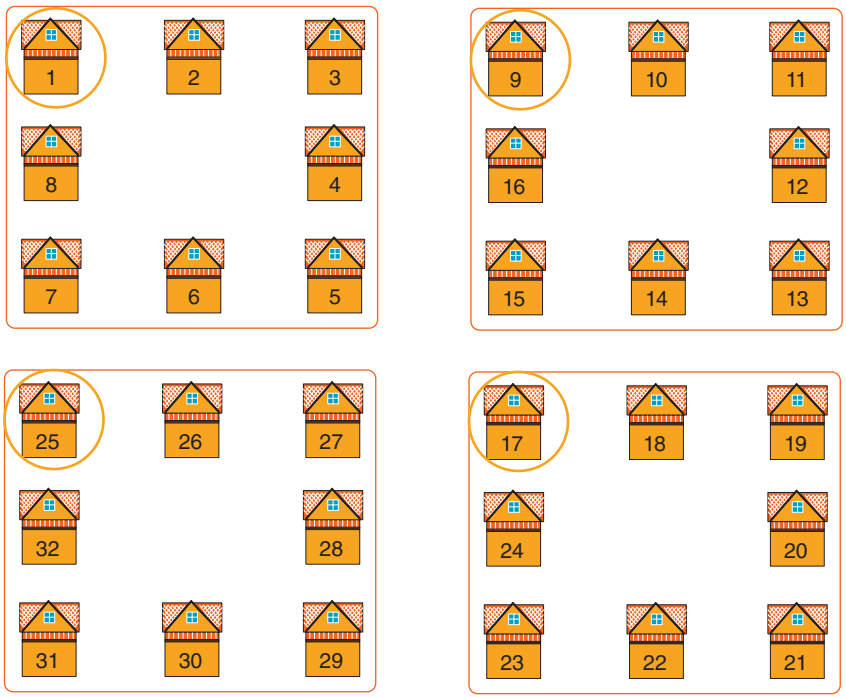
Systematic random sampling is a variant of simple random sampling. The first element is selected randomly from a list or from sequential files, and then every *n*th element is selected. This is a convenient method for drawing a random sample when the population elements are arranged sequentially. It is particularly efficient when the elements are not actually available in a list (i.e., there is no sampling frame) but instead are represented by folders in filing cabinets. Systematic random sampling is also useful for sampling from a mobile population, such as visitors to a museum during some period (Kalton 2020:19). For example, every tenth visitor who enters the front door could be invited to participate in a brief survey.

Systematic random sampling requires the following three steps:

1. The total number of cases in the population is divided by the number of cases required for the sample. This division yields the sampling interval, the number of cases from one sampled case to another. If 50 cases are to be selected out of 1,000, the **sampling interval** is 20; every 20th case is selected.
2. A number from 1 to 20 (or whatever the sampling interval is) is selected randomly. This number identifies the first case to be sampled, counting from the first case on the list or in the files.
3. After the first case is selected, every *n*th case is selected for the sample, where *n* is the sampling interval. If the sampling interval is not a whole number, the size of the sampling interval is varied systematically to yield the proper number of cases for the sample. For example, if the sampling interval is 30.5, the sampling interval alternates

between 30 and 31. In almost all sampling situations, systematic random sampling yields what is essentially a simple random sample. The exception is a situation in which the sequence of elements is affected by **periodicity**—that is, the sequence varies in some regular, periodic pattern. For example, the houses in a new development with the same number of houses in each block (e.g., 8) may be listed by block, starting with the house in the northwest corner of each block and continuing clockwise. If the sampling interval is 8, the same as the periodic pattern, all the cases selected will be in the same position (see Figure 5.6). If houses on those corner lots are different from others (perhaps the lots are larger), then the sample will be biased. But in reality, periodicity and the sampling interval are rarely the same.

FIGURE 5.6 ■ The Effect of Periodicity on Systematic Random Sampling



If the sampling interval is 8 for a study in this neighborhood, every element of the sample will be a house on the northwest corner—and thus the sample will be biased.

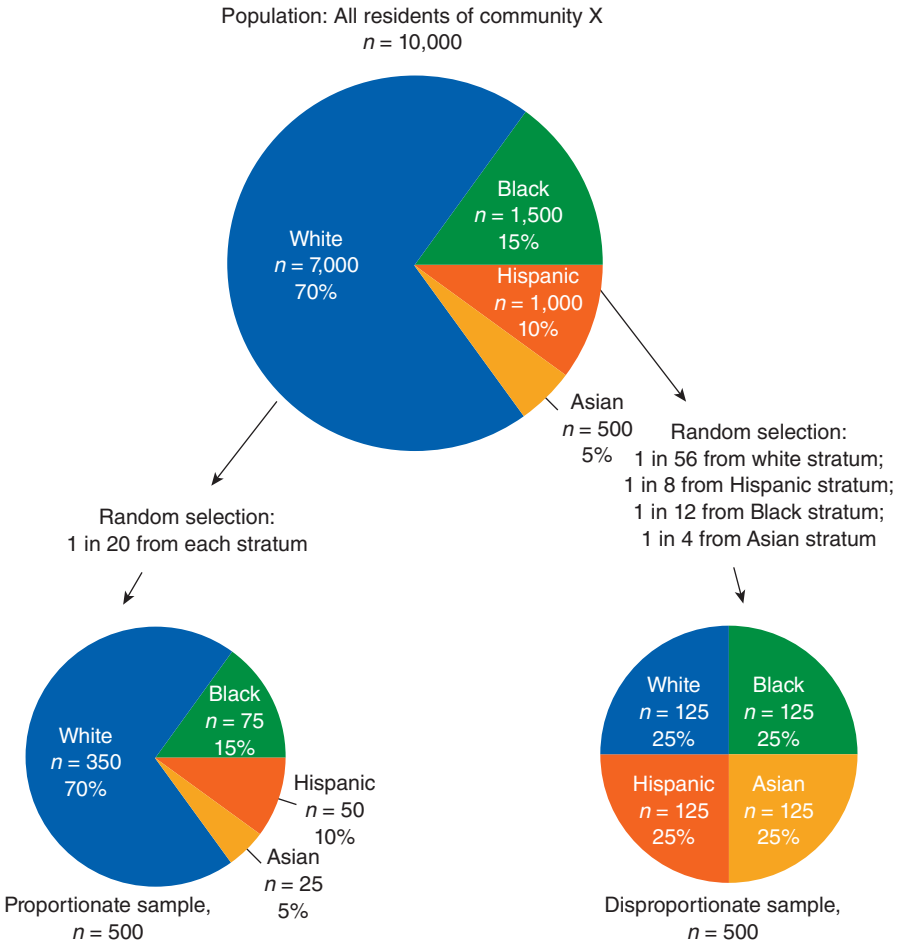
Stratified Random Sampling

Although all probability sampling methods use random sampling, some add steps to the sampling process to make sampling more efficient or easier. **Stratified random sampling** uses information known about the total population before sampling to make the sampling process more efficient. First, all elements in the population (i.e., in the sampling frame) are distinguished according to their value on some relevant characteristic. This characteristic might be year in school in a study of students, marital status in a study of family relations, or average property value in a study of towns. That characteristic forms the sampling strata. Next, elements are sampled randomly from within these strata. For example, race may be the basis for

distinguishing individuals in some population of interest. Within each racial category, individuals are then sampled randomly. Of course, using this method requires more information before sampling than is the case with simple random sampling. It must be possible to categorize each element in one and only one stratum, and the size of each stratum in the population must be known.

This method is more efficient than drawing a simple random sample because it ensures appropriate representation of elements across strata (Kalton 2020:23). Imagine you plan to draw a sample of 500 from the population of a large company to study the experiences of different ethnic groups. You know from company records that the workforce is 15% Black, 10% Hispanic, 5% Asian, and 70% white. If you drew a simple random sample, you might end up with somewhat disproportionate numbers of each group. But if you created sampling strata based on race and ethnicity, you could randomly select the correct proportion of cases from each stratum: 75 Blacks (15% of the sample), 50 Hispanics (10%), 25 Asians (5%), and 350 whites (70%). By using **proportionate stratified sampling**, you would eliminate any possibility of sampling error in the sample's distribution of ethnicity. Each stratum would be represented exactly in proportion to its size in the population from which the sample was drawn (see Figure 5.7).

FIGURE 5.7 ■ Stratified Random Sampling

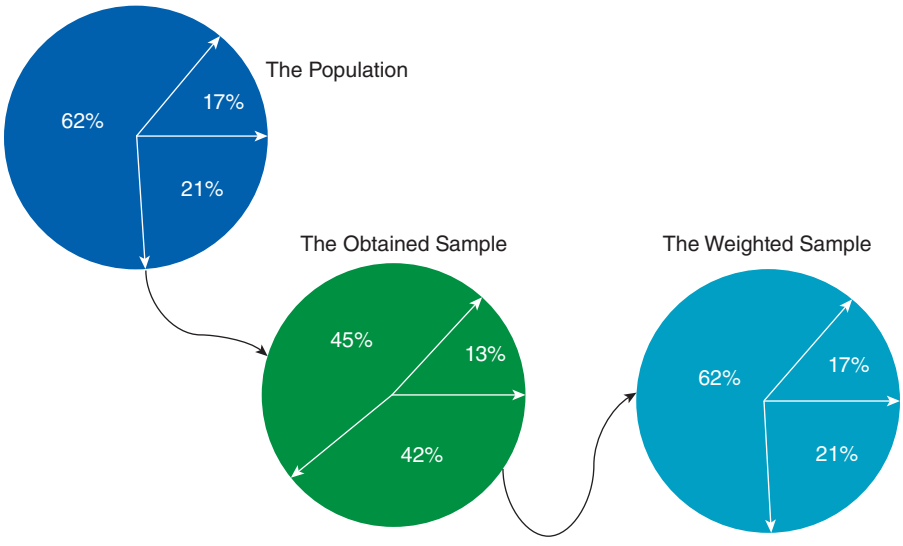


In **disproportionate stratified sampling**, the proportion of each stratum that is included in the sample is intentionally varied from what it is in the population. In the case of the company sample stratified by ethnicity, you might select equal numbers of cases from each racial or ethnic group: 125 Blacks (25% of the sample), 125 Hispanics (25%), 125 Asians (25%), and 125 whites (25%). In this type of sample, the probability of selection of every case is known but unequal between strata. You know what the proportions are in the population, and so you can easily adjust your combined sample statistics to reflect these true proportions. For instance, if you want to combine the ethnic groups and estimate the average income of the total population, you will have to *weight* each case in the sample. The weight is a number you multiply by the value of each case based on the stratum it is in. For example, you would multiply the incomes of all Black persons in the sample by 0.6 (75/125), the incomes of all Hispanic persons by 0.4 (50/125), and so on. Weighting in this way reduces the influence of the oversampled strata and increases the influence of the under-sampled strata to what they would have been if pure probability sampling had been used.

Why would anyone select a sample that is unrepresentative in the first place? The most common reason is to ensure that cases from smaller strata are included in the sample in sufficient numbers to allow separate statistical estimates and to facilitate comparisons between strata (Kalton 2020:28). Remember one of the determinants of sample quality is sample size. The same is true for subgroups within samples. If a key concern in a research project is to describe and compare the incomes of people from different racial and ethnic groups, then it is important that the researchers base the mean income of each group on enough cases to be a valid representation. If few members of a particular underrepresented group are in the population, they need to be oversampled. Such disproportionate sampling may also result in a more efficient sampling design if the costs of data collection differ markedly between the strata or if the variability (heterogeneity) of the strata differs.

Weighting is also sometimes used to reduce the lack of representativeness of a sample that occurs because of nonresponse. On finding that the obtained sample does not represent the population for some known characteristics, such as, perhaps, gender or education, the researcher weights the cases in the sample so the sample has the same proportions of men and women or high school graduates and college graduates as the complete population (see Figure 5.8).

FIGURE 5.8 ■ Weighting an Obtained Sample to Match a Population Proportion



Keep in mind, though, this procedure does not solve completely the problems caused by an unrepresentative sample because you still don't know what the sample composition should have been relative to the other variables in your study; all you have done is to reduce the sample's unrepresentativeness relative to the variables used in weighting. This weighting may, in turn, make it more likely that the sample is representative of the population relative to other characteristics, but you don't really know. The underestimation of support for President Trump in the 2020 presidential polls, discussed earlier in this chapter, is an example of this limitation of weighting: Although pollsters weighted their samples to compensate for lower response rates from less educated whites in 2020 (they had learned the importance of doing that in 2016), support for Trump was higher among those in this category who didn't respond than among those who participated in polls, so giving more weight to less educated white respondents still did not lead to an accurate estimate of the degree of support for Trump.

The University of Massachusetts Boston Center for Survey Research designed a stratified sampling strategy for our National Science Foundation-funded study, *Coronavirus and Urban Neighborhoods*. A table based on the resulting methodology report illustrates the process and logic of disproportionate random sampling (see Table 5.1). Figures from 8 of the 25 neighborhoods sampled are included in the table.

In collaboration with the Mayor's Office and Northeastern University's Boston Area Research Initiative (BARI), Boston was first divided into 25 neighborhoods that reflected traditional urban divisions, could be aligned with Census block groups, and were used by BARI to organize its "big data" resources on city activities and characteristics (see chapter 14). The 25 neighborhoods were treated as the sampling strata and household addresses were then randomly selected within each neighborhood. The number of addresses selected in each neighborhood varied with the total neighborhood population to result in a sufficient number of cases for making neighborhood-specific population estimates, even in smaller neighborhoods. In other words, the probability of selection was higher in less populous neighborhoods. In addition, we oversampled in four neighborhoods with higher percentages of Black and/or Hispanic residents to ensure enough survey respondents to permit analyses focused specifically on these groups.

TABLE 5.1 ■ NSF-Survey Neighborhood Sampling Specifications and Response

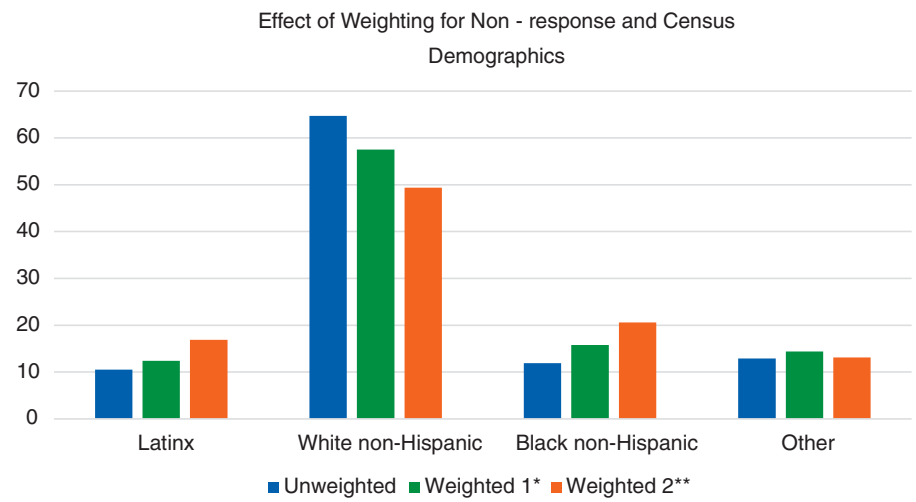
Neighborhood	*N* Sampled Addresses	Probability of Selection	*N* Completed Surveys	Response Rate
Hyde Park	364	.02967	59	17.10%
West Roxbury	189	.01407	59	32.24%
Roslindale	188	.01820	73	40.11%
Mattapan	362	.02704	61	17.58%
Dorchester Central	189	.01042	39	21.08%
Dorchester South	191	.01671	60	32.97%
Dorchester North	188	.02661	42	23.86%
Lower Roxbury	372	.05977	57	17.59%

Source: Roman, Tony. 2020. *NSF Beacon Survey: Methodology.* Boston: Center for Survey Research, University of Massachusetts. Unpublished report.

The initial mailing contained the eight-page questionnaire, a postage paid return envelope, a fact sheet, and a cover letter about the survey and a $2 cash incentive. Questionnaires sent to three neighborhoods with high percentages of Hispanic households received surveys printed in both English and Spanish. After two weeks, a similar mailing was sent again to nonrespondents (although without another $2 incentive). These procedures resulted in similar numbers of surveys from most neighborhoods (including those that were oversampled, such as Hyde Park, Mattapan, and Lower Roxbury), but with neighborhood response rates that varied from 17.10% to 40.11% (26.88% for the entire city).

Estimates from the survey sample to the entire population of the City of Boston had to consider the differential probabilities of selection of households in different neighborhoods, as well as the differences in the response rates between different groups. This was accomplished through calculating a weight for each respondent. The result was that the weighted sample matched the distribution of city residents across neighborhoods as well as the characteristics of Boston's entire population in terms of age, gender, race/ethnicity, and education (as reported by the U.S. Bureau of the Census). Figure 5.9 illustrates the effect of weighting on the distribution of race/Hispanic origin for the citywide sample. The final column percentages match perfectly the distribution of race/Hispanic origin in the Census Bureau's American Community Survey data for Boston.

FIGURE 5.9 ■ Unweighted and Weighted Sample Distribution of Race/Hispanic Origin



*Weighted for probability of selection and nonresponse by neighborhood

**Weighted also to match citywide distribution by age, gender, race/ethnicity, age (American Community Survey, Census Bureau, 2014-2018 average).

Source: Coronavirus and Urban Neighborhoods, original data.

How does this sample strike you? Let's assess sample quality using the questions posed earlier in the chapter:

- *From what population were the cases selected?* There is a clearly defined population: the adult residents of the City of Boston (who live in households, rather than in group quarters like college dorms, prisons, or shelters).

- *What method was used to select cases from this population?* The case selection method is a random selection procedure, using disproportionate stratified sampling by neighborhood, with no systematic biases in the sampling.
- *Do the cases that were studied represent, in the aggregate, the population from which they were selected?* The findings will very likely represent the population sampled because there were no biases in the sampling and a very large number of cases were selected. However, only about 27% of those selected for the survey responded, so we do not know the attitudes of those who did not respond. Even after adjustment for demographic characteristics, this rate of nonresponse reduces confidence in the sample's representativeness of the attitudes and behaviors measured in the survey.

We must also consider the issue of cross-population generalizability: Do findings from this sample have implications for any larger group beyond the population from which the sample was selected? Because a representative sample of the entire U.S. adult population was not drawn, this question has to do with generalizations to other cities or to suburban and rural areas. The many differences between cities, states, and types of communities in responses to the coronavirus pandemic make it clear that cross-population generalizations should be made with caution. Generalizations to other nations would require very careful consideration.

Multistage Cluster Sampling

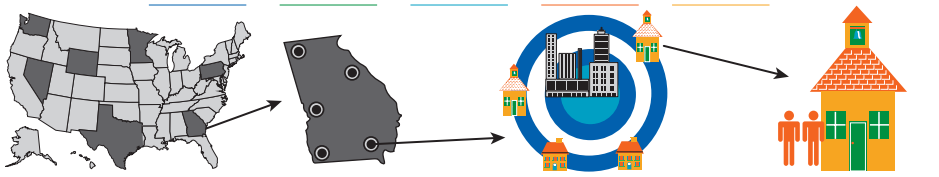
Multistage cluster sampling is useful when a sampling frame of elements is not available, as often is the case for large populations spread out across a wide geographic area or among many different organizations (Kalton 2020:3). A **cluster** is a naturally occurring, mixed aggregate of elements of the population, with each element appearing in one, and only one, cluster. Schools could serve as clusters for sampling students, blocks could serve as clusters for sampling city residents, counties could serve as clusters for sampling the general population, and businesses could serve as clusters for sampling employees.

Drawing a multistage cluster sample is, at least, a two-stage procedure. First, the researcher draws a random sample of clusters. A list of clusters should be much easier to obtain than a list of all the individuals in each cluster in the population. Next, the researcher draws a random sample of elements within each selected cluster. Because only a fraction of the total clusters is involved, obtaining the sampling frame at this stage should be much easier.

In a cluster sample of city residents, for example, blocks could be the first-stage clusters. A research assistant could walk around each selected block and record the addresses of all occupied dwelling units. Or in a cluster sample of students, a researcher could contact the schools selected in the first stage and plan with the registrar to obtain lists of students at each school. Cluster samples often involve more than two stages (see Figure 5.10), with clusters within clusters, as when a national sample of individuals might involve first sampling states, then geographic units within those states, then dwellings within those units, and finally, individuals within the dwellings. In multistage cluster sampling, the clusters at the first stage of sampling are termed the *primary sampling units* (Levy and Lemeshow 1999:228).

How many clusters should be selected, and how many individuals within each cluster should be selected? Generally, the sample will be more similar to the entire population if the researcher selects as many clusters as possible—even though this will mean the selection of fewer individuals within each cluster. Unfortunately, this strategy also maximizes the cost of the sample for studies using in-person interviews. The more clusters a researcher selects, the more time and money will have to be spent traveling to the different clusters to reach the individuals for interviews.

FIGURE 5.10 ■ Multistage Cluster Sampling



Stage 1: Randomly select states

Stage 2: Randomly select cities, towns, and counties within those states

Stage 3: Randomly select schools within those cities and towns

Stage 4: Randomly select students within each school

The calculation of how many clusters to sample and how many individuals are within the clusters is also affected by the degree of similarity of individuals within clusters: The more similar the individuals are within the clusters, the fewer the number of individuals needed to represent each cluster. However, the selected clusters are more likely to represent the entire population of clusters if they are more internally heterogeneous (Kalton 2020:36). So if you set out to draw a cluster sample, be sure to consider how similar individuals are within the clusters as well as how many clusters you can afford to include in your sample.

Multistage cluster sampling is a very popular method among survey researchers, but it has one general drawback: Sampling error is greater in a cluster sample than in a simple random sample because there are two steps involving random selection rather than just one. This sampling error increases as the number of clusters decreases, and it decreases as the homogeneity of cases per cluster increases. In sum, it's better to include as many clusters as possible in a sample, and it's more likely that a cluster sample will be representative of the population if cases are relatively similar within clusters.

Probability Sampling Methods Compared

Can you now see why researchers often prefer to draw a stratified random sample or a cluster sample rather than a simple random sample? Table 5.2 should help you remember the key features of these different types of sampling and to determine when each is most appropriate.

TABLE 5.2 ■ Features of Probability Sampling Methods

Feature	Simple	Systematic	Stratified	Cluster
Unbiased selection of cases	Yes	Yes	Yes	Yes
Sampling frame required	Yes	No	Yes	No
Ensures representation of key strata	No	No	Yes	No
Uses natural grouping of cases	No	No	No	Yes
Reduces sampling costs	No	No	No	Yes
Sampling error compared with simple random sample	–	Same	Lower	Higher

Many professionally designed surveys use combinations of cluster and stratified probability sampling methods. For example, the General Social Survey (GSS) is an ongoing biennial (annually from 1972 to 1993) social survey project funded by the National Science Foundation and administered by the National Opinion Research Center (NORC) to track social characteristics and attitudes in the Untied States. New cross-sectional samples are selected for each survey. Because it relies on in-person interviews, the GSS achieves a high rate of response that only began to drop below 70% since 2012 (Morgan 2020).

The population for the GSS is the adult households of the United States.—not including persons living in group quarters like dorms, nursing homes, or prisons (2.6% of U.S. adults in the 2010 U.S. Census). Those who do not speak English or (after 2004) Spanish are also excluded. Since there is no national directory of all adult U.S. residents from which a sample can be selected directly, the GSS uses a multistage cluster design, with also some stratification within strata. The first stage includes Metropolitan Statistical Areas (MSAs) and other counties; the second stage includes census tracts and block groups, while the third stage samples blocks, from which households are then sampled. One adult is randomly selected from all eligible adults within a household (Marsden and Smith 2016).

How would we evaluate the GSS samples, using the sample evaluation questions?

- *From what population were the cases selected?* The population is clearly defined as adults in the United States living in households and speaking English or Spanish.
- *What method was used to select cases from this population?* The random selection method is carefully described.
- *Do the cases that were studied represent, in the aggregate, the population from which they were selected?* The unbiased selection procedures and high rate of response create confidence in the representativeness of the sample. In addition, special efforts are made to reach a sample of nonrespondents, so their characteristics can be estimated. A recent comparison to population characteristics identified by the Census Bureau's American Community Survey also indicates that the demographic characteristics of the GSS sample are very similar to those of the U.S. adult population (Morgan 2020).

NONPROBABILITY SAMPLING METHODS

Nonprobability sampling methods are often used in qualitative research; they also are used in quantitative studies when researchers are unable to use probability selection methods. In qualitative research, a focus on one setting or a very small sample allows a more intensive portrait of activities and actors and may lead to confidence in having represented fairly the entire setting and participants in it. Even when this argument seems persuasive, though, we must still be concerned with the external validity of the findings—the extent to which they can be generalized to other places. In the many studies in which qualitative researchers have interviewed only a small portion of the participants in a setting under investigation or observed just a limited portion of the events of interest, we have to realize that the use of nonprobability sampling methods limits the ability to generalize to the whole population of participants or events in the setting and lowers the confidence that others can place in these generalizations.

There are four common nonprobability sampling methods: (1) availability sampling, (2) quota sampling, (3) purposive sampling, and (4) snowball sampling. Because these methods do not use a random selection procedure, we cannot expect a sample selected with any of these methods to yield a representative sample. They should not be used in quantitative studies if a probability-based method is feasible. Nonetheless, these methods are useful when random sampling is not possible, when a research question calls for an intensive investigation of a small population, or when a researcher is performing a preliminary, exploratory study.

Availability (Convenience) Sampling

Elements are selected for **availability sampling**—also called convenience sampling, haphazard sampling, or accidental sampling—because they're available or easy to find. There are many ways to select elements for an availability sample: standing on street corners and talking to whoever walks by, asking questions of employees who have time to talk when they pick up their paychecks at a personnel office, or approaching particular individuals at opportune times while observing activities in a social setting. You may find yourself interviewing available students at campus hangouts as part of a course assignment. A participant observation study of a group may require no more sophisticated approach.

Megan A. Carney (2017) used availability sampling in her study of food insecurity among Mexican and Central American migrant women. Focusing on Santa Barbara County in California, she worked with local service organizations and community health workers to publicize her study and recruit potential interviewees. Women were eligible for participation if they were at least 18 years old, had migrated from Mexico or Central America, and had used some form of food assistance in the United States. Over a 20-month period between 2008 and 2001, Carney interviewed 25 women selected in this way.

Now, answer the sample evaluation questions in relation to Carney's migrant women interviewees. If your answers are something like "The population was intended to be migrant women from Mexico and Central America in Santa Barbara who had used food assistance," "The method for selecting cases was based on availability," and "We cannot determine the extent to which these women studied represent the population," you're right! There certainly is not much likelihood in such an availability sample that the interviewees represented the distribution of experiences among all migrant women in Santa Barbara. However, if you have read much about migrant women from similar areas and have come to believe there are many more similarities than differences between them, you might believe their experiences of food insecurity are likely to be similar across settings.

An availability sample is appropriate in social research when a field researcher is exploring a new setting and trying to get some sense of the prevailing attitudes or when a survey researcher conducts a preliminary test of a new set of questions, but it can also be required when the population simply cannot be listed or otherwise identified in advance—at least with available resources. It was because of the lack of an adequate sampling frame (list of the population) and sufficient resources that Kristine et al. (2021) used an availability sampling strategy in their study of HIV testing among Muslim women in the United States. In an increasingly common approach to developing a sample, Hearld and her colleagues recruited American Muslim women for their online survey through online social networks, e-mail, and announcements on existing websites involving Muslim groups. Interested individuals were directed to an online website with more information about the study and the link to the survey. Two hundred and eighteen Muslim women completed the survey.

As with the Carney study, we do not know how well the responses of these 218 women represent those that would be obtained from a survey of the entire eligible population. The people who happen to be available in any situation are unlikely to be just like those who are unavailable. We can't be at all certain that what we learn can be generalized with any confidence to a larger population of concern. Nonetheless, analysis of data collected with an availability sample can lead to valuable insights about people's feelings or behaviors (as happened in the Carney and Hearld studies).

Availability sampling often masquerades as a more rigorous form of research. Popular magazines periodically survey their readers by printing a questionnaire for readers to fill out and mail in. A follow-up article then appears in the magazine under a title such as "What You Think About Intimacy in Marriage." If the magazine's circulation is large, a large sample can be achieved in this way. The problem is that usually only a tiny fraction of readers return the questionnaire, and these respondents are probably unlike other readers who did not have the interest in participating or the time to do so. Even though the follow-up article may be interesting, we have no basis for thinking the results based on this availability sample describe the readership as a whole—much less the population at large.

Do you see now why availability sampling differs so much from random sampling methods, which require that "nothing but chance" affects the actual selection of cases? What makes availability sampling "haphazard" is precisely that a great many things other than chance can affect the selection of cases, ranging from possible prejudices of the research staff to the work schedules of potential respondents. To truly leave the selection of cases up to chance, we have to design the selection process very carefully so other factors are not influential. There's nothing haphazard about selecting cases randomly.

Quota Sampling

Quota sampling is intended to overcome the most obvious flaw of availability sampling—that the sample will just consist of whoever or whatever is available, without any concern for its similarity to the population of interest. The distinguishing feature of a quota sample is that quotas are set to ensure that the sample represents certain characteristics in proportion to their prevalence in the population.

Suppose you want to sample adult residents of a town in a study of support for a tax increase to improve the town's schools. You know from the town's annual report what the proportions of town residents are in gender, race, age, and number of children. You think each of these characteristics might influence support for new school taxes, so you want to be sure the sample includes men; women; persons who are white, Black, Hispanic, Asian; older people; younger people; big families; small families; and childless families in proportion to their numbers in the town population.

This is where quotas come in. Let's say that 48% of the town's adult residents are men and 52% are women; that 60% are employed; 5% are unemployed, and 35% are out of the labor force. These percentages and the percentages corresponding to the other characteristics become the quotas for the sample. If you plan to include 500 residents in your sample, 240 must be men (48% of 500), 260 must be women, 300 must be employed, and so on (see Table 5.3).

You may even set more refined quotas, such as certain numbers of employed women, employed men, unemployed men, and so on. With the quota list in hand, you (or your research staff) can now go out into the community looking for the right number of people in each quota category. You may go door to door or bar to bar or just stand on a street corner until you have surveyed 240 men, 260 women, and so on.

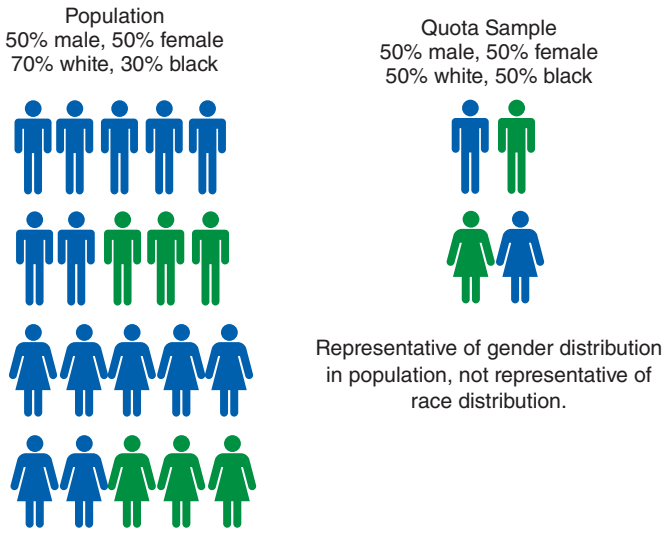
TABLE 5.3 ■ Adjusting a Quota Sample for Population

Characteristic	Population	Quota	Calculation
Sex			
Men	48%	240	(.48*500)
Women	52%	260	(.52*500)
Employment Status			
Employed	60%	300	(.60*500)
Unemployed	5%	25	(.05*500)
Out of Labor Force	35%	175	(.35*500)
Total Sample Size	500		

The problem is that even when we know that a quota sample is representative of the characteristics for which quotas have been set, we have no way of knowing if the sample is representative for any other characteristics. In Figure 5.11, for example, quotas have been set for gender only. Under these circumstances, it's no surprise that the sample is representative of the population only for gender, not race. Interviewers are only human; they may avoid potential respondents with menacing dogs in the front yard, or they could seek out respondents who are physically attractive or who look like they'd be easy to interview. Quota sampling replaces people who are unavailable or unwilling to participate with someone else, so it will increase the sample's bias in favor of more cooperative people (Kalton 2020:139). Realistically, researchers can set quotas for only a small fraction of the characteristics relevant to a study, so a quota sample may not be much better than an availability sample (although following careful, consistent procedures for selecting cases within the quota limits always helps).

This last point leads me to another limitation of quota sampling: You must know the characteristics of the entire population to set the right quotas. In most cases, researchers know what

FIGURE 5.11 ■ Quota Sampling


Population
50% male, 50% female
70% white, 30% black
Quota Sample
50% male, 50% female
50% white, 50% black
Representative of gender distribution in population, not representative of race distribution.

the population looks like relative to no more than a few of the characteristics relevant to their concerns—and in some cases, they have no such information on the entire population.

If you're now feeling skeptical of quota sampling, you've gotten the drift of my remarks.

Nonetheless, in some situations, establishing quotas can add rigor to sampling procedures. It's almost always better to maximize possibilities for comparison in research, and quota sampling techniques can help qualitative researchers do this. For instance, (Timmer et al. 1993:7) interviewed homeless persons in several cities and other locations for their book on the sources of homelessness. Persons who were available were interviewed, but the researchers paid some attention to generating a diverse sample. They interviewed 20 homeless men who lived on the streets without shelter and 20 mothers who were found in family shelters. About half of those whom the researchers selected in the street sample were Black, and about half were white. Although the researchers did not use quotas to try to match the distribution of characteristics among the total homeless population, their informal quotas helped ensure some diversity in key characteristics.

Does quota sampling remind you of stratified sampling? It's easy to understand why because they both select sample members partly based on one or more key characteristics. Table 5.4 identifies the differences between quota sampling and stratified random sampling. The key difference, of course, is the lack of random selection in quota sampling.

TABLE 5.4 ■ Comparison of Stratified and Quota Sampling Methods

Feature	Stratified	Quota
Unbiased (random) selection of cases	Yes	No
Sampling frame required	Yes	No
Ensures representation of key strata	Yes	Yes

Purposive Sampling

In **purposive sampling**, each sample element is selected for a purpose, usually because of the unique position of the sample elements. Purposive sampling may involve studying the entire population of some limited group (such as directors of shelters for immigrants) or a subset of a population (perhaps mid-level managers with a reputation for efficiency). Or a purposive sample may be a *key informant survey*, which targets individuals who are particularly knowledgeable about the issues under investigation.

Herbert Rubin and Irene Rubin (1995) suggest three guidelines for selecting informants when designing any purposive sampling strategy. Informants should be

- Knowledgeable about the cultural arena or situation or experience being studied
- Willing to talk
- Represent[ative of] the range of points of view (p. 66)

In addition, Rubin and Rubin (1995) suggest continuing to select interviewees until you can pass two tests:

- *Completeness:* What you hear provides an overall sense of the meaning of a concept, theme, or process. (p. 72)
- *Saturation:* You gain confidence that you are learning little that is new from subsequent interview[s]. (p. 73)

Adhering to these guidelines will help ensure that a purposive sample adequately represents the setting or issues studied.

Of course, purposive sampling does not produce a sample that represents some larger population, but it can be exactly what is needed in a case study of an organization, community, or some other clearly defined and relatively limited group. In an intensive organizational case study, a purposive sample of organizational leaders might be complemented with a probability sample of organizational members. Before designing her probability samples of hospital patients and homeless persons, Dee Roth (1990:146–147) interviewed a purposive sample of 164 key informants from organizations that had contact with homeless people in each of the counties she studied.

Snowball Sampling

Snowball sampling is useful for hard-to-reach or hard-to-identify populations for which there is no sampling frame but the members of which are somewhat interconnected (at least some members of the population know each other). It can be used to sample members of groups such as drug dealers, sex workers, practicing criminals, participants in Alcoholics Anonymous groups, gang leaders, informal organizational leaders, and homeless persons. It may also be used for charting the relationships between members of some group (a sociometric study), for exploring the population of interest before developing a formal sampling plan, and for developing what becomes a census of informal leaders of small organizations or communities.

Kristine Hearld and her colleagues (2021:19) added a snowball sampling component to their study of HIV testing among Muslim women that primarily used availability sampling. For the snowball sampling component, potential respondents who went to the study website were in turn asked to share the survey with other American Muslim women they knew. However, researchers using snowball sampling cannot be confident that their sample represents the total population of interest, in part because individuals with larger social networks are more likely to be included. Generalizations must be tentative.

Respondent-Driven Sampling

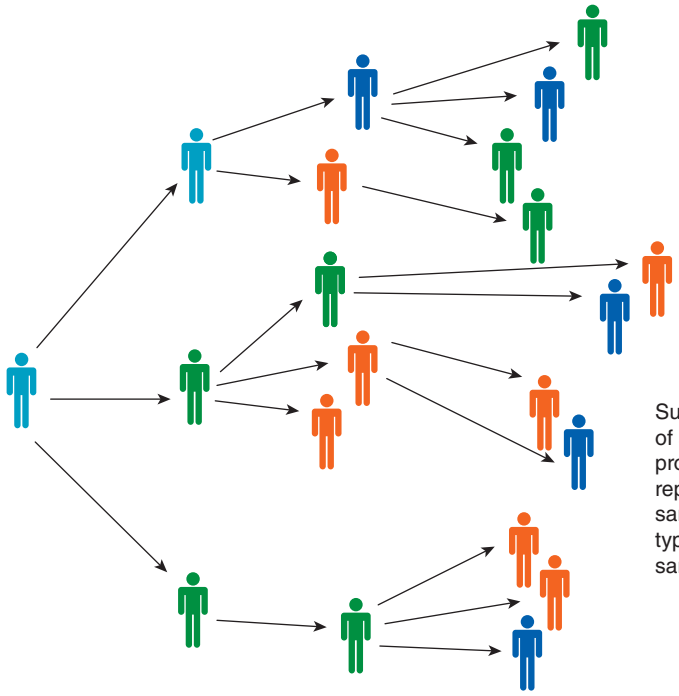
A more systematic version of snowball sampling can reduce the potential for bias. *Respondent-driven sampling* (RDS) gives financial incentives to respondents to recruit diverse peers (Heckathorn 1997; Khoury 2020). Limitations on the number of incentives any one respondent can receive increase the sample's diversity. Targeted incentives can steer the sample to include specific subgroups. When the sampling is repeated through several waves, with new respondents bringing in more peers, the composition of the sample converges on a more representative mix of characteristics than would occur with uncontrolled snowball sampling. Figure 5.12 shows how the sample spreads out through successive recruitment waves to an increasingly diverse pool (Heckathorn 1997:178). Figure 5.13 shows that even if the starting point were all white persons, respondent-driven sampling would result in an appropriate ethnic mix from an ethnically diverse population (Heckathorn 2002:17). However, it is often difficult to carry out an RDS sampling strategy as planned, so generalizability of findings from the resulting sample should be evaluated carefully (Kalton 2020:143).

Lessons About Sample Quality

Some lessons are implicit in my evaluations of the samples in this chapter:

- We can't evaluate the quality of a sample if we don't know what population it is supposed to represent. If the population is unspecified because the researchers were

FIGURE 5.12 ■ Respondent-Driven Sampling

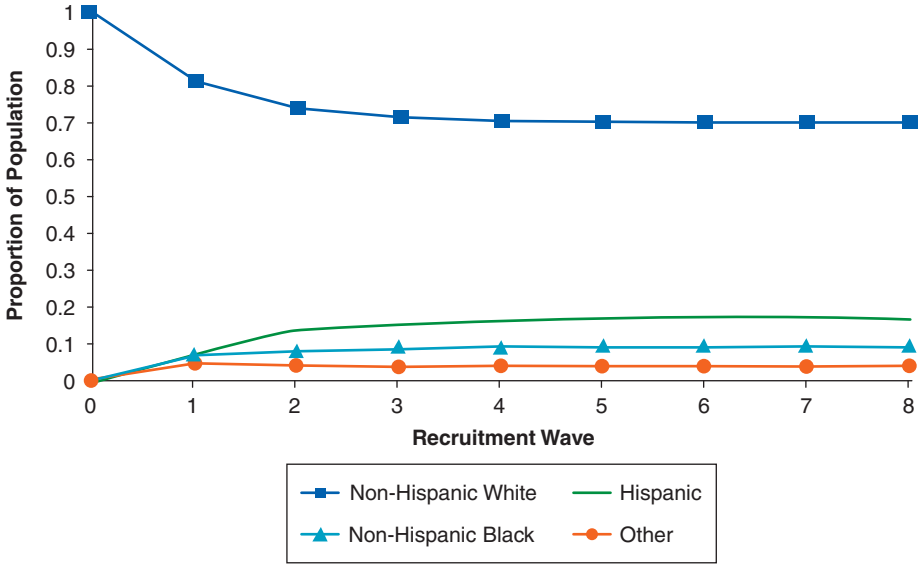


Successive waves of sampling gradually produce a more representative sample than is typical of snowball sampling.

Instructions to respondents:
"We'll pay you $5 each for up to three names, but only one of those names can be somebody from your own town. The others have to be from somewhere else."

Source: Based on Heckathorn (1997), p. 178.

FIGURE 5.13 ■ Convergence of Respondent-Driven Sample to True Ethnic Proportions in Population, After Starting With Only Whites



Source: Copyright © 2002, The Society for the Study of Social Problems Inc. Reprinted with permission from Oxford University Press.

never clear about the population they were trying to sample, then we can suspect that the sample itself has major problems.

- We can't evaluate the quality of a sample if we don't know how cases in the sample were selected from the population. If the method was specified, we then need to know whether cases were selected in a systematic fashion and on the basis of chance. In any case, we know that a haphazard method of sampling (as in person-on-the-street interviews) undermines generalizability.
- Sample quality is determined by the sample obtained, not just by the sampling method itself. If many of the people selected for our sample do not respond or do not participate in the study for some other reason, although they have been selected for the sample, the quality of our sample is undermined—even if we chose the sample in the best possible way.
- We need to be aware that even researchers who obtain very good samples may talk about the implications of their findings for some group that is larger than or just different from the population they sampled. For example, findings from a representative sample of students in one university often are discussed as if they tell us about university students in general. And maybe they do; we just don't know for sure.
- A sample that allows comparisons involving theoretically important variables is often better than one that does not allow such comparisons. Even when we study people or social processes in depth, it is best to select individuals or settings with an eye to how useful they will be for examining relationships. Limiting an investigation to just one setting or just one type of person will inevitably leave us wondering what it is that makes a difference.

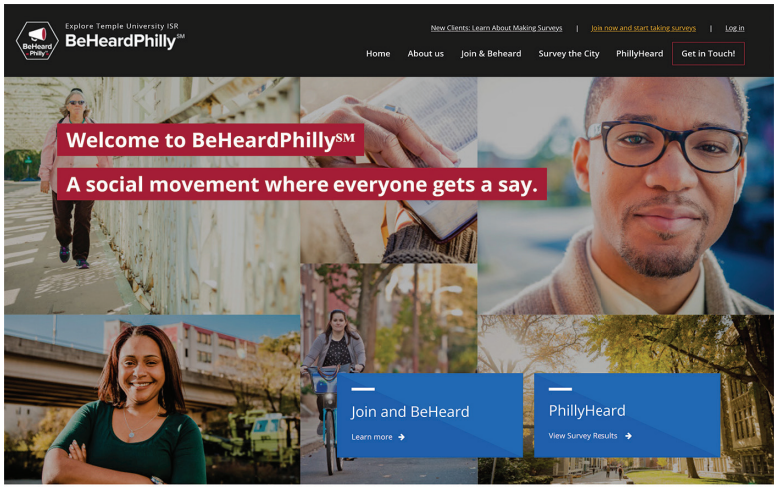
Generalizability With Web-Based Samples

As web-based social interaction has increased—dramatically during the coronavirus pandemic—social researchers have also turned to the Internet to study samples from populations of interest. Chapter 8 reviews different approaches to conducting surveys on the web, but it's important first to consider the challenges of sampling on the Internet in the context of generalizing to larger populations.

One popular approach to research on the web is to post an "opt-in" survey on the web. Whether the survey is aimed at the population of a city, nation, or the world, an attractively designed welcome page and survey instrument can result in many thousands of interested people. For example, the *BeHeardPhilly* program designed by Temple University's Institute for Social Research and funded in part by city agencies and nonprofits has attracted thousands of Philadelphia residents to answer questions designed to help inform program planning in areas like housing, transportation, and health (see Photo 5.2). It is a remarkable effort to give residents a direct voice about social issues of concern in their neighborhoods.

But who chooses to opt-in and complete a survey on the web about city concerns? First, only a small fraction of the total population will participate in surveys like this, even with extensive publicity and some financial incentives for participating. In 2019, there were 3,695 participants enrolled in the BeHeardPhilly panel out of a total adult population of Philadelphia of more than 1.2 million (2020 U.S. Census data https://data.census.gov/cedsci/profile?g=0600000US4210160000; City of Philadelphia and Institute for Social Research 2021:40). That's .29% of the population who opted in. When these panelists were asked to

complete a special survey in 2019, only 23.7% of that .29% participated. Second, those who choose to participate tend to be much more highly educated and more often white than the population from which the sample is obtained (City of Philadelphia and Institute for Survey Research 2021:42–43). Of course, those in the tiny sliver of the population that chooses to participate are likely to differ in many other ways from those who don't participate. As a result, as in other availability sampling methods, findings cannot be generalized with confidence to the population (Kalton 2020:143).



BeHeardPhilly homepage

https://beheardphilly.com/

The U.S. Census Bureau used a more ambitious approach to obtaining a representative sample for its Household Pulse Survey (HPS), designed to measure the impact of the coronavirus pandemic. Rather than just opening its 20-minute online survey to anyone, the Census Bureau sent an invitation to participate in the survey by e-mail or text message to a representative sample of households. Although the invitation informed potential respondents that the survey was legally authorized, respondents were not legally required to complete the survey (unlike the actual census).

In just one week at the end of 2021 (Phase 3.3, Week 40), 1,044,103 households were invited to participate in the HPS; 5.8% responded. The Census Bureau then weighted these 60,826 participants so their demographic characteristics matched the distribution of age, sex, race, and Hispanic origin as reported in Census Bureau statistics for their state (U.S. Bureau of the Census n.d.c). However, even after equating these demographic characteristics of the survey respondents with the actual distributions for the populations in their state, we do not know whether the survey respondents' answers to questions about the impact of the pandemic represented the sentiments of the approximately 95% of the population who did not respond to the HPS.

Generalizability in Qualitative Research

Qualitative research often focuses on populations that are hard to locate or very limited in size. In consequence, nonprobability sampling methods such as availability sampling and snowball sampling are often used. However, this does not mean that generalizability should be ignored in qualitative research or that a sample should be studied simply because it is convenient (Gobo

2008:206). Janet Ward Schofield (2002) suggests two different ways of increasing the generalizability of the samples obtained in such situations:

> *Studying the Typical.* Choosing sites on the basis of their fit with a typical situation is far preferable to choosing on the basis of convenience. (p. 181)
>
> *Performing Multisite Studies.* A finding emerging repeatedly in the study of numerous sites would appear to be more likely to be a good working hypothesis about some as yet unstudied site than a finding emerging from just one or two sites. . . . Generally speaking, a finding emerging from the study of several very heterogeneous sites would be more . . . likely to be useful in understanding various other sites than one emerging from the study of several very similar sites. (p. 184)

Giampietro Gobo (2008:204–205) highlights another approach to improving generalizability in qualitative research. A case may be selected for in-depth study because it is atypical, or deviant. Investigating social processes in a situation that differs from the norm will improve understanding of how social processes work in typical situations: "the exception that proves the rule."

Rather than attempting to improve generalizability, some qualitative researchers instead question its value as a goal. The argument is that understanding the particulars of a situation in depth is an important object of inquiry in itself. In the words of sociologist Norman Denzin,

> The interpretivist rejects generalization as a goal and never aims to draw randomly selected samples of human experience. . . . Every instance of social interaction . . . represents a slice from the life world that is the proper subject matter for interpretive inquiry. (Denzin cited in Schofield 2002:173)

But see if you can read the results of an intensive study of even a single setting or group without considering the extent to which the findings can be generalized to other settings or groups. I don't think it's possible.

SAMPLING DISTRIBUTIONS

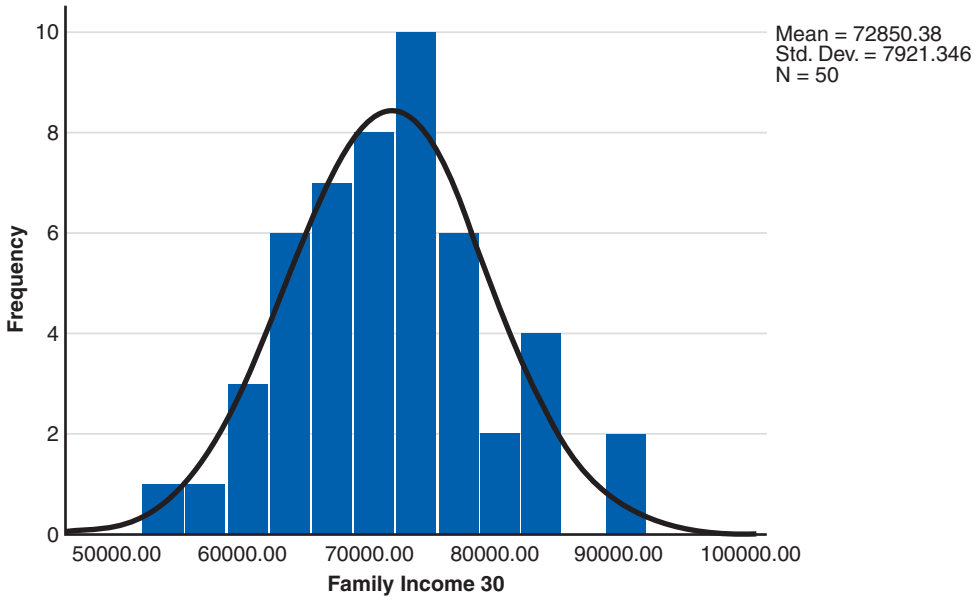
A well-designed probability sample is one that is likely to be representative of the population from which it was selected. But as you've seen, random samples still are subject to sampling error owing just to chance. To deal with that problem, social researchers consider the properties of a sampling distribution, a hypothetical distribution of a statistic across all the random samples that could be drawn from a population. Any single random sample can be thought of as just one of an infinite number of random samples that, in theory, could have been selected from the population. If we had the finances of Gatsby and the patience of Job and were able to draw an infinite number of samples and we calculated the same type of statistic for each of these samples, we would then have a true sampling distribution. Understanding sampling distributions is the foundation for understanding how statisticians can estimate sampling error.

What does a sampling distribution look like? Because a sampling distribution is based on some statistic calculated for different samples, we need to choose a statistic. Let's focus on the arithmetic average or mean. I will explain the calculation of the mean in Chapter 9, but you may already be familiar with it: You add up the values of all the cases and divide by the total number of cases. Let's say you draw a random sample of 500 families from the entire population of a state and find their average (mean) family income is $58,239. Imagine you then draw another

random sample. That sample's mean family income might be $60,302. Imagine marking these two means on graph paper and then drawing more random samples and marking their means on the graph. The resulting graph would be a sampling distribution of the mean.

Figure 5.14 demonstrates what happened when I did something very similar to what I have just described—not with an infinite number of samples and not from a large population but through the same process using the 2020 General Social Survey (GSS) sample as if it were a population. First, I drew 50 different random samples, each consisting of 30 cases, from the 2020 GSS. (The standard notation for the number of cases in each sample is n = 30.) Then I calculated for each random sample the approximate mean family income (approximate because the GSS does not record actual income in dollars). I then graphed the means of the 50 samples. Each bar in Figure 5.14 shows how many samples had a mean family income in categories about $3000 wide. The mean for the "population" (the total GSS sample in this example) is $73,529, and you can see that many of the samples in the sampling distribution are close to this value, with the mean for this sampling distribution of 50 samples being $72,850—very close to the "population" mean. However, although many of the sample means are close to the population mean, some are quite far from it (the lowest is actually $53,691, while the highest is $90,183). If you had calculated the mean from only one sample, it could have been anywhere in this sampling distribution, but it is *unlikely* to have been far from the population mean—that is, unlikely to have been close to either end (or "tail") of the distribution. Most of the sample means in the sampling distribution are closer to the population mean.

FIGURE 5.14 ■ Partial Sampling Distribution: Mean Family Income (Samples of Size 30) Output Generated From IBM SPSS Statistics, Version 28.



Source: General Social Survey 2020 (National Opinion Research Center [NORC] 2021).

Estimating Sampling Error

We don't observe actual sampling distributions in real research; researchers just draw the best sample they can and then are stuck with the results—one sample, not a distribution of samples. A sampling distribution is a theoretical distribution. However, we can use the properties of

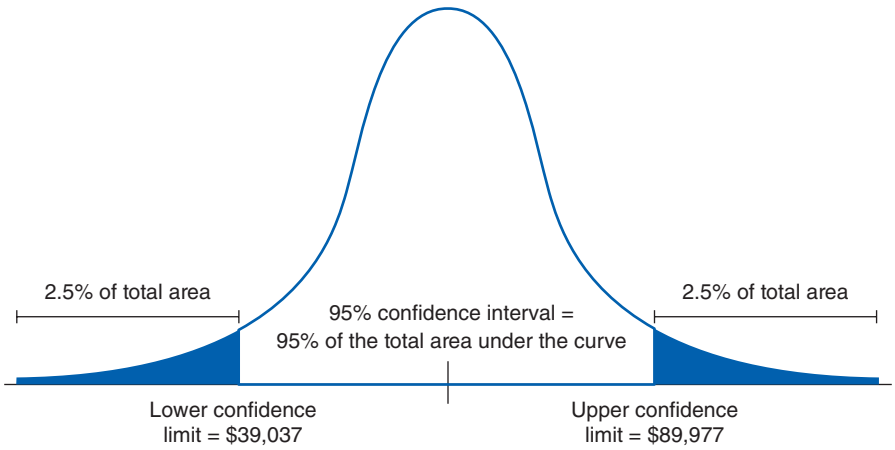
sampling distributions to calculate the amount of sampling error that was likely with the random sample used in a study. The tool for calculating sampling error is called **inferential statistics**.

Sampling distributions for many statistics, including the mean, have a "normal" shape. A graph of a normal distribution looks like a bell, with one hump in the middle, centered on the population mean, and the number of cases tapering off on both sides of the mean. Note that a normal distribution is symmetric: If you folded it in half at its center (at the population mean), the two halves would match perfectly. This shape is produced by **random sampling error**—variation owing purely to chance. The value of the statistic varies from sample to sample because of chance, so higher and lower values are equally likely.

The partial sampling distribution in Figure 5.14 does not have a completely normal shape because it involves only a small number of samples (50), each of which has only 30 cases. The normal distribution superimposed over the bars in Figure 5.14 shows how close the distribution of 50 samples is to a perfectly normal distribution.

Figure 5.15 shows some of the properties of a true normal distribution—how a sampling distribution like the one in Figure 5.14 would appear if, rather than 50 random samples, I had selected thousands of random samples.

FIGURE 5.15 ■ Normal Sampling Distribution: Mean Family Income


2.5% of total area
95% confidence interval =
95% of the total area under the curve
2.5% of total area
Lower confidence limit = $39,037
Upper confidence limit = $89,977
Mean Family Income = $64,507

The properties of a sampling distribution facilitate the process of statistical inference. In the sampling distribution, the most frequent value of the **sample statistic**—the statistic (such as the mean) computed from sample data—is identical to the **population parameter**—the statistic computed for the entire population. In other words, we can have a lot of confidence that the value at the peak of the bell curve represents the norm for the entire population. A population parameter also may be termed the *true value* for the statistic in that population. A sample statistic is an estimate of a population parameter.

In a normal distribution, a predictable proportion of cases fall within certain ranges. Inferential statistics takes advantage of this feature and allows researchers to estimate how likely it is that, given a particular sample, the true population value will be within some range of the statistic. For example, a statistician might conclude from a sample of 30 families, "We can be 95% confident that the true mean family income in the total population is between $39,037 and $89,977." The interval from $39,037 to $89,977 would then be called the *95% confidence interval for the mean.* The lower ($39,037) and upper ($89,977) bounds of this interval are termed

the *confidence limits*. Figure 5.15 marks such confidence limits, indicating the range that encompasses 95% of the area under the normal curve; 95% of all sample means would fall within this range, as does the mean of our hypothetical sample of 30 cases.

Although all normal distributions have these same basic features, they differ from one another in the extent to which they cluster around the mean. A sampling distribution is more compact when it is based on larger samples. Stated another way, we can be more confident in estimates based on larger random samples because we know that a larger sample creates a more compact sampling distribution. A sampling distribution based on samples of 100 randomly selected cases each would be much tighter, with shorter tails, compared to the sampling distribution of samples of 30 cases each in Figure 5.14. By contrast, a sampling distribution based on samples of 5 randomly selected cases each would be much wider, with longer tails, compared to the sampling distribution in Figure 5.14. With very small samples, the confidence intervals around our statistics become very wide, indicating that we can have very little confidence in precise estimates from the sample to the population.

Other confidence intervals, such as the 99% confidence interval, can be reported. As a matter of convention, statisticians use only the 95%, 99%, and 99.9% confidence limits to estimate the range of values that are likely to contain the true value. These conventional limits reflect the conservatism inherent in classical statistical inference: Don't make an inferential statement about the population unless you are very confident (at least 95% confident) that it is correct. In any case, you should now have a sense of how researchers make inferences from a random sample of a population.

Sample Size Considerations

You have learned now that more confidence can be placed in the generalizability of statistics from larger samples, so that you may be eager to work with random samples that are as large as possible. Unfortunately, researchers often cannot afford to sample a very large number of cases. Therefore, they try to determine during the design phase of their study how large a sample they must have to achieve their purposes. They have to consider the degree of confidence desired, the homogeneity of the population, the complexity of the analysis they plan, and the expected strength of the relationships they will measure.

- The less sampling error desired, the larger the sample size must be.
- Samples of more homogeneous populations can be smaller than can samples of more diverse populations. Stratified sampling uses prior information on the population to create more homogeneous population strata from which the sample can be selected, so stratified samples can be smaller than simple random samples of the same population.
- If the only analysis planned for a survey sample is to describe the population in terms of a few variables, a smaller sample is required than if a more complex analysis involving sample subgroups is planned. If much of the analysis will focus on estimating the characteristics of subgroups within the sample, the size of the subgroups must be considered rather than the size of the total sample (Levy and Lemeshow 1999:74).
- When the researchers expect to find very strong relationships between the variables when they test hypotheses, they will need a smaller sample to detect these relationships than if they expect weaker relationships.

Researchers can make more precise estimates of the sample size required through a method called *statistical power analysis* (Kraemer and Thiemann 1987). Statistical power analysis requires

a good advance estimate of the strength of the hypothesized relationship in the population. In addition, the math is complicated, so it helps to have some background in mathematics or to be able to consult a statistician. For these reasons, many researchers do not conduct formal power analyses when deciding how many cases to sample.

You can obtain some general guidance about sample sizes from the current practices of social scientists. For professional studies of the national population in which only a simple description is desired, professional social science studies typically have used a sample size of between 1,000 and 1,500 people, with as many as 2,500 being included if detailed analyses are planned. Studies of local or regional populations often sample only a few hundred people, partly because these studies lack sufficient funding to draw larger samples. Of course, the sampling error in these smaller studies is considerably larger than in a typical national study (Sudman 1976:87).

CONCLUSIONS

Sampling is a powerful tool for social science research. Probability sampling methods allow a researcher to use the laws of chance, or probability, to draw samples from which population parameters can be estimated with a high degree of confidence. A sample of just 1,000 or 1,500 individuals can be used to estimate reliably characteristics of the population of a nation comprising millions of individuals.

But researchers do not come by representative samples easily. Well-designed samples require careful planning, some advance knowledge about the population to be sampled, and adherence to systematic selection procedures—all so that the selection procedures are not biased. And even after the sample data are collected, the researcher's ability to generalize from the sample findings to the population is not completely certain. The best they can do is to perform additional calculations that state the degree of confidence that can be placed in the sample statistic.

The alternatives to random, or probability-based, sampling methods are almost always much less palatable for quantitative studies, even though they are typically much cheaper. Without a method of selecting cases likely to represent the population in which the researcher is interested, research findings will have to be carefully qualified. Qualitative researchers whose goal is to understand a small group or setting in depth may necessarily have to use unrepresentative samples, but they must keep in mind that the generalizability of their findings will not be known. Additional procedures for sampling in qualitative studies are introduced in Chapter 10.

Social scientists often seek to generalize their conclusions from the population that they studied to some larger target population. The external validity of generalizations of this type is necessarily uncertain because having a representative sample of a particular population does not at all ensure that what we find will hold true in other populations. Nonetheless, as you will see in Chapter 16, the cumulative findings from studies based on local or otherwise unrepresentative populations can provide important information about broader populations.

KEY TERMS

Availability sampling (p. 175)
Census (p. 157)
Cluster (p. 172)
Disproportionate stratified sampling (p. 169)
Elements (p. 152)
Enumeration units (p. 153)
Inferential statistics (p. 185)
Multistage cluster sampling (p. 172)

Nonprobability sampling method (p. 158)
Nonrespondents (p. 160)
Periodicity (p. 167)
Population (p. 152)
Population parameter (p. 185)
Probability of selection (p. 158)
Probability sampling method (p. 158)
Proportionate stratified sampling (p. 168)
Purposive sampling (p. 178)
Quota sampling (p. 176)
Random digit dialing (p. 165)
Random number table (p. 165)
Random sampling (p. 159)
Random sampling error (chance sampling error) (p. 185)
Replacement sampling (p. 165)
Representative sample (p. 158)
Sample (p. 152)
Sample statistic (p. 185)
Sampling error (p. 155)
Sampling frame (p. 152)
Sampling interval (p. 166)
Sampling units (p. 153)
Simple random sampling (p. 165)
Snowball sampling (p. 179)
Stratified random sampling (p. 167)
Systematic bias (p. 160)
Systematic random sampling (p. 166)
Target population (p. 154)

HIGHLIGHTS

LO 5.1 Summarize the process of probability-based sampling.

- The purpose of sampling is to be able to generalize what you have learned from the sample to the population. Specifically, we would like to be able to generalize to our target population a set of elements to which the researcher would ideally like to generalize any study findings, but that may differ from the actual population studied due to practical or other constraints.
- To generate a sample that gives you a valid picture of your population, you select elements (individual members of the population whose characteristics are to be measured) from a sampling frame (a list of all elements or other units containing the elements in a population).
- In studies in which the individuals or other entities from which we collect information are not actually the elements, a multistage sampling design is used. Sampling units are the units listed at each stage of a multistage sampling design. Primary sampling units are the units selected in the first stage of sampling. Secondary sampling units are the units selected in the second stage of sampling. For example, a researcher might sample schools (primary sampling units) for a survey about educational practices and then interview a sample of teachers (secondary sampling units) in each sampled school to obtain the information about educational practices.

LO 5.2 Describe the circumstances that make sampling unnecessary and the reasons they are rare.

- To assess sample quality and whether sampling is necessary, when you are planning or evaluating a study, ask yourself these questions: From what population were the cases selected? What method was used to select cases from this population? Do the cases that were studied represent, in the aggregate, the population from which they were selected?

- Sampling is unnecessary when the elements that would be sampled are identical, but the complexity of the social world makes it difficult to argue very often that all different elements are identical.
- Conducting a complete census of a population also eliminates the need for sampling, but the resources required for a complete census of a large population are usually prohibitive.

LO 5.3 Explain why probability-based sampling is needed to obtain a representative sample.

- Probability sampling methods allow us to know in advance how likely it is that any element of a population will be selected for the sample—their "probability of selection."
- Probability sampling methods rely on a random, or chance, selection procedure, which is, in principle, the same as flipping a coin to decide which of two people "wins" and which one "loses."
- Using probability sampling methods eliminates systematic bias in the selection of elements for the sample.
- An unbiased (random) sample is a representative sample, and findings from the sample can be generalized to the population from which the sample was selected.
- A random sample will still have some sampling error resulting from chance. The likely degree of sampling error can be calculated for a random sample.
- The larger the sample, the more confidence we can have in the sample's representativeness.
- The more homogeneous the population, the more confidence we can have in the representativeness of a sample of a particular size.
- The fraction of the total population that a sample contains does not affect the sample's representativeness unless that fraction is large.
- Nonresponse can result in systematic bias in a sample that was designed for probability-based selection of cases.

LO 5.4 Define the different probability sampling methods.

There are four common methods for drawing random samples: (1) simple random sampling, (2) systematic random sampling, (3) stratified random sampling, and (4) multistage cluster sampling:

- Simple random sampling is sampling in which every sample element is selected only based on chance, through a random process. The random selection of cases can use a random number table (Appendix B), a lottery-like procedure in which small cards or other objects are mixed up prior to selection, or a computer program.
- Systematic random sampling is a variant of simple random sampling. The first element is selected randomly from a list or from sequential files, and then every *n*th element is selected. The sampling interval is the number of cases from one sampled case to another in a systematic random sample.
- Stratified random sampling uses information known about the total population before sampling to make the sampling process more efficient. First, all elements in the population (i.e., in the sampling frame) are distinguished according to their value on some relevant characteristic. That characteristic forms the sampling strata. Next, elements are sampled randomly from within these strata. Using this

method requires more information before sampling than is the case with simple random sampling. It must be possible to categorize each element in one and only one stratum, and the size of each stratum in the population must be known.

- Multistage cluster sampling is a sampling method in which elements are selected in two or more stages, with the first stage being the random selection of naturally occurring clusters and the last stage being the random selection of elements within clusters. Multistage cluster sampling is useful when a sampling frame of elements is not available, as often is the case for large populations spread out across a wide geographic area or among many different organizations (Kalton 2020:3).

LO 5.5 Defend the value of nonprobability sampling methods.

- Nonprobability sampling methods can be useful when random sampling is not possible, when a research question does not concern a larger population, and when a preliminary exploratory study is appropriate. However, the representativeness of nonprobability samples cannot be determined.
- There are four common nonprobability sampling methods: (1) availability sampling, (2) quota sampling, (3) purposive sampling, and (4) snowball sampling:
  - Availability sampling (also called convenience sampling, haphazard sampling, or accidental sampling) involves the selection of elements because they're available or easy to find. An availability sample is appropriate in social research when a field researcher is exploring a new setting and trying to get some sense of the prevailing attitudes or when a survey researcher conducts a preliminary test of a new set of questions, but it can also be required when the population simply cannot be listed or otherwise identified in advance—at least with available resources
  - Quota sampling involves setting quotas to ensure that the sample represents certain characteristics in proportion to their prevalence in the population. In some situations, establishing quotas can add rigor to sampling procedures. It's almost always better to maximize possibilities for comparison in research, and quota sampling techniques can help qualitative researchers do this.
  - Purposive sampling involves sample elements being selected for a purpose, usually because of the unique position of the sample elements. Purposive sampling does not produce a sample that represents some larger population, but it can be exactly what is needed in a case study of an organization, community, or some other clearly defined and relatively limited group
  - Snowball sampling involves the selection of sample elements as they are identified by successive informants or interviewees. Snowball sampling is useful for hard-to-reach or hard-to-identify populations for which there is no sampling frame but the members of which are somewhat interconnected (at least some members of the population know each other). Respondent-driven sampling is a more systematic approach to snowball sampling that uses recruitment incentives to obtain a more diverse sample.

LO 5.6 Assess how sampling error is affected by sample size, the heterogeneity of the population, and the fraction of the population included in the sample.

- The likely degree of error in an estimate of a population characteristic based on a probability sample decreases when the size of the sample increases.

- The likely degree of error in an estimate of a population characteristic based on a probability sample decreases when the homogeneity of the population from which the sample was selected increases.
- The proportion of the population that is sampled does not affect sampling error, except when that proportion is large.
- The degree of sampling error affecting a sample statistic can be estimated from the characteristics of the sample and knowledge of the properties of sampling distributions.

DISCUSSION QUESTIONS

1. When (if ever) is it reasonable to assume that a sample is not needed because "everyone is the same"—that is, the population is homogeneous? Does this apply to research such as that of Stanley Milgram on obedience to authority? What about investigations of student substance abuse? How about investigations of how people (or their bodies) react to alcohol? What about research on the likelihood of voting (the focus of Chapter 9)?

2. All adult U.S. citizens are required to participate in the decennial census, but some do not. Some social scientists have argued for putting more resources into a large representative sample, so that more resources are available to secure higher rates of response from hard-to-include groups. Do you think that the U.S. Census should shift to a probability-based sampling design? Why or why not?

3. What increases sampling error in probability-based sampling designs? Stratified rather than simple random sampling? Disproportionate (rather than proportionate) stratified random sampling? Stratified rather than cluster random sampling? Why do researchers select *disproportionate* (rather than proportionate) stratified samples? Why do they select cluster rather than simple random samples?

4. What are the advantages and disadvantages of probability-based sampling designs compared with nonprobability-based designs? Could any of the research described in this chapter with a nonprobability-based design have been conducted instead with a probability-based design? What are the difficulties that might have been encountered in an attempt to use random selection? How would you discuss the degree of confidence you can place in the results obtained from research using a nonprobability-based sampling design?

5. How much pressure is too much pressure to participate in a probability-based sample survey? Is it okay for the U.S. government to mandate legally that all citizens participate in the decennial census? Should companies be able to require employees to participate in survey research about work-related issues? Should students be required to participate in surveys about teacher performance? Should parental consent be required to allow their high school–age students to participate in a survey at school about substance abuse and health issues? Is it okay to give monetary incentives for participation in a survey of homeless shelter clients? Can monetary incentives be coercive? Explain your decisions.

SPSS EXERCISES

1. Take a look at the distribution of support for capital punishment (cappun_2) with what is called a *frequency distribution.*
   a. Click Analyze/Descriptive Statistics/Frequencies.
   b. Highlight cappun_2 and click on the arrow that sends it over to the Variables window, then click OK.

      Examine the percentages in the Valid percent column. What percentage of the U.S. population as represented in the GSS2020 favored capital punishment?

2. Now select random samples of the GSS2020 respondents:
   a. Go to the Data Editor window, and select a random sample containing 40 of the respondents. From the menu:
      i. Click Data/Select cases/Random sample of cases.
      ii. Choose Sample.
      iii. For "Sample Size," choose Exactly and then enter 40 cases from the first 100 cases.
      iv. Click Continue/OK. (Before you click OK, be sure that the "Filter out unselected cases" box is checked.)
   b. Determine the percentage of the subsample that favored capital punishment by repeating the steps in SPSS Exercise 1. Record the subsample characteristics and its percentage.
   c. Now, repeat the steps in 2a 10 times. Each time, add 100 to the "first 100 cases" request (so that on the last step you will be requesting "Exactly 40 cases from the first 1,000 cases").
   d. Select a random sample containing five of the respondents. Now repeat the steps in 2a 10 times, this time for samples of 5.

3. How does the distribution of cappun_2 in these subsamples compare with that for the total GSS sample?
   a. Plot the percents produced in Step 2c and 2d on separate sheets of graph paper. Each graph's horizontal axis will represent the possible range of percentages (from 0 to 100, perhaps in increments of 5); the vertical axis will represent the number of samples in each range of percentages (perhaps ranging from 0 to 10). Make an *X* to indicate this percentage for each sample. If two samples have the same percentage, place the corresponding Xs on top of each other. The X for each sample should be one unit high on the vertical axis.
   b. Draw a vertical line corresponding to the point on the horizontal axis that indicates the percentage of the total GSS sample that favors capital punishment.
   c. Describe the shape of both graphs. These are the sampling distributions for the two sets of samples. Compare them with each other. Do the percentages from the larger samples tend to be closer to the mean of the entire sample (as obtained in SPSS Exercise 1)? What does this tell you about the relationship between sample size and sampling error?

©iStockphoto.com/South_agency

6 RESEARCH DESIGN AND CAUSATION

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

6.1 Explain individual and group units of analysis and the role they play in the ecological and reductionist fallacies

6.2 Distinguish the three different longitudinal designs

6.3 Illustrate an example of a nomothetic and of an idiographic causal explanation

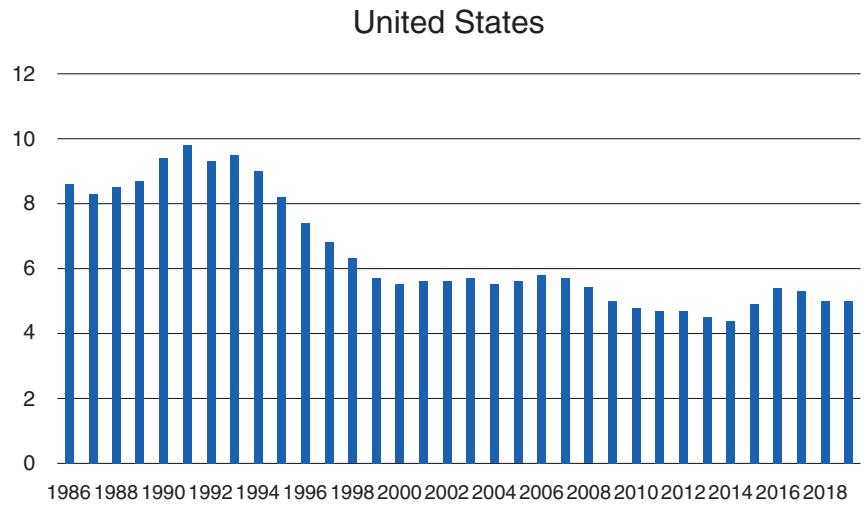
6.4 List the three criteria for establishing a nomothetic causal relationship

6.5 Recommend two cautions a researcher should keep in mind to improve understanding of a causal connection

6.6 Appraise the strengths and weaknesses of employing an experimental design as compared to other methods

Identifying causes—figuring out why things happen—is the goal of most social science research. Unfortunately, valid explanations of the causes of social phenomena do not come easily. Why did the number and rate of homicides rise in the early 1990s and then begin a sustained drop that leveled off in the 2000s, then declined a bit more, then rose somewhat in the mid-2010s (Smith and Cooper 2013; FBI 2021) (see Figure 6.1)? Arizona State University criminologist Scott Decker points to the low levels of crime committed by undocumented immigrants to explain the falling crime rate in his state (Archibold 2010), whereas criminal justice advocates in Texas point to the state's investment in community treatment and diversion programs (Grissom 2011). Police officials in New York City point to the effectiveness of CompStat, a computer program that highlights the location of crimes (Dewan 2004a:A25; Dewan, 2004b:A1; Kaplan, 2002:A3), but others think New York City has benefited from its

FIGURE 6.1 ■ Rate of Homicides in the United States, 1986–2019



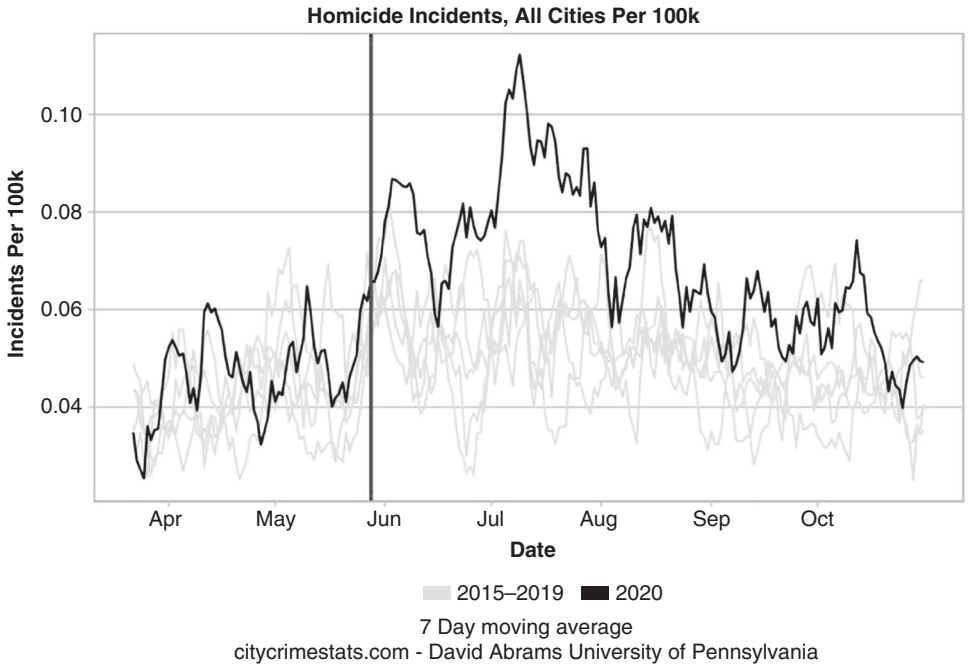
Source: Federal Bureau of Investigation. 2021. "Trend of Homicide from 1986 to 2019." *Crime Data Explorer,* September 18 (https://crime-data-explorer.app.cloud.gov/pages/explorer/crime/crime-trend).

Safe Streets, Safe Cities program (Rashbaum 2002) or from a decline in crack cocaine use (Dewan 2004b:C16).

But then the picture changed during the pandemic. You can see in Figure 6.2 that the homicide rate began to climb after the pandemic shutdown began in March 2020, and then accelerated sharply two months later—much more than in the summer months in the previous four years. Was this due to economic turmoil and widespread unemployment? But other types of crime didn't increase. Was it the effect of domestic abuse victims trapped at home (Southall 2020)? Maybe hesitancy of residents to turn to the police for help after unrest following George Floyd's murder by a Minneapolis police officer (MacFarquhar 2021A)? Enrique Cardiel, community organizer and public health worker in Albuquerque, concluded that "people are desperate and they don't have a lot of options, so they turn toward violence as a way to solve things" (MacFarquhar 2021B:A19). Albuquerque Police Department Chief Harold Medina felt "it is a perfect storm, ...not just one factor that we can point to," but included consequences of COVID-19 and the social justice protests (MacFarquhar 2021B:A1). How about the closure of recreation centers, group therapy programs, summer jobs programs, violence reduction programs (Badger and Bui 2020)? What features of research designs can help us answer questions like these?

I begin this chapter by discussing three key elements of research design: the design's units of analysis, its use of cross-sectional or longitudinal data, and whether its methods are primarily quantitative or qualitative. Whenever we design research, we must decide whether to use individuals or groups as our units of analysis and whether to collect data at one or several points in time. The decisions that we make about these design elements will affect our ability to draw causal conclusions in our analysis. Whether

FIGURE 6.2 ■ Homicides Before and After Stay-At-Home Orders, 37-City Average



Source: Abrams, D. S. 2020. COVID and Crime: Key Findings; citycrimestats.com.

the design is primarily quantitative or qualitative in its methods also affects the type of causal explanation that can be developed: Quantitative projects lead to *nomothetic* causal explanations, whereas qualitative projects that have a causal focus can lead to *idiographic* explanations (described in the section "Quantitative or Qualitative Causal Explanations"). After reviewing these three key design elements, I will also review the criteria for achieving explanations that are causally valid from a nomothetic perspective. By the end of the chapter, you should have a good grasp of the different meanings of causation and be able to ask the right questions to determine whether causal inferences are likely to be valid. You should also have a broader understanding of the causes of crime and violence.

UNITS OF ANALYSIS

In nonexperimental research designs, we can be misled about the existence of an association between two variables when we do not know to what **units of analysis** the measures in our study refer—that is, the level of social life on which the research question is focused, such as individuals, groups, towns, or nations. I first discuss this important concept before explaining how it can affect causal conclusions.

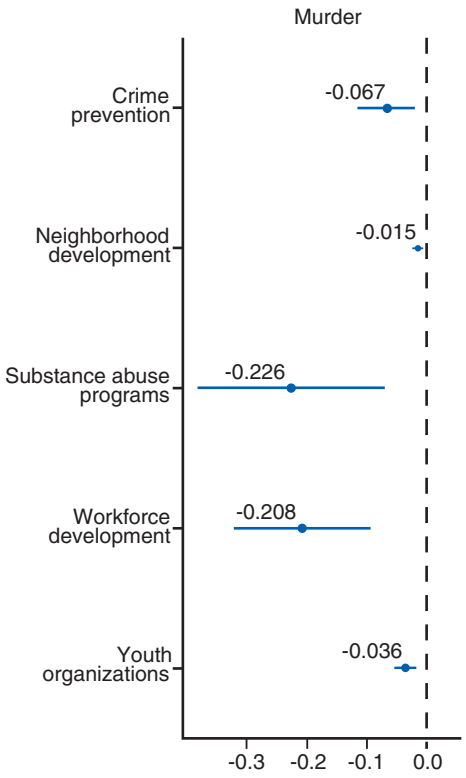
Individual and Group

In most sociological and psychological studies, the units of analysis are individuals. The researcher may collect survey data from individuals, analyze the data, and then report on, say, how many experienced domestic violence during the pandemic and whether financial stress predicted a greater risk of violence. This is what Australian researchers Anthony Morgan and Hayley Boxall (2020) studied with a sample of 7,514 women in a cohabiting relationship who had responded to an online survey of Australian women 18 years or older in May 2020. They found the risk of physical or sexual violence from a partner increased for women reporting more financial stress due to the pandemic, but only for those who had not experienced physical or sexual violence from their partner before February 2020. The risk of physical or sexual violence did not change for women who had previously experienced domestic violence due to heightened financial stress during the pandemic.

The units of analysis may instead be groups of some sort, such as families, schools, work organizations, towns, states, or countries. For example, Patrick Sharkey, Gerard Torrats-Espinosa, and Delaram Takyar (2017) hypothesized that an increase in the number of community nonprofits would reduce the rate of homicides and other crimes in cities. For their analysis, Sharkey et al (2017). used group-level data on number of nonprofits and crime rates from 264 cities over more than 20 years. Figure 6.3 shows cities that added more community nonprofits—particularly substance abuse programs and workforce development programs—experienced a greater decline in their murder rate (the same was true for other forms of violent crime and nonviolent property crimes). Because the data describe the cities, cities are the (group) units of analysis.

In some studies, groups are the units of analysis, but data are collected from individuals. For example, in their study of influences on violent crime in Chicago neighborhoods, Robert Sampson et al. (1997:919) hypothesized that collective efficacy would influence neighborhood crime rates. *Collective efficacy* was defined conceptually as a characteristic of the neighborhood: the extent to which residents were likely to help other residents and were trusted by other residents. However, Sampson et al. (1997) measured this variable in a survey of individuals. The

FIGURE 6.3 ■ Estimates of Effects on Crime Rate of Different Types of Community Nonprofits



Source: Sharkey, Patrick, Gerard Torrats-Espinosa, and Delaram Takyar. 2017. "Community and the Crime Decline: The Causal Effect of Local Nonprofits on Violent Crime." *American Sociological Review,* 82(6):1231.

responses of individual residents about their perceptions of their neighbors' helpfulness and trustworthiness were averaged together to create a collective efficacy score for each neighborhood. This neighborhood measure of collective efficacy was used to explain variation in the rate of violent crime between neighborhoods. The data were collected from individuals and were about individuals, but they were combined (aggregated) to describe neighborhoods. The units of analysis were thus groups (neighborhoods).

In a study such as that of Sampson, Raudenbush, and Earls (1997), we can distinguish the concept of units of analysis from the **units of observation**. Data were collected from individuals, the units of observation in this study, and then the data were aggregated and analyzed at the group level.

The important point is to know what the units of observation are and what the level of analysis is and then to evaluate whether the conclusions are appropriate to these study features. A conclusion that "crime increases with joblessness" could imply either that individuals who lose their jobs are more likely to commit a crime or that a community with a high unemployment rate is likely to have a high crime rate—or both. We also must know what the units of analysis are to interpret statistics appropriately. Measures of association tend to be stronger for group-level than for individual-level data because measurement errors at the individual level tend to cancel out at the group level (Bridges and Weis 1989:29–31).

The Ecological Fallacy and Reductionism

Researchers should make sure their causal conclusions reflect the units of analysis in their study. Conclusions about processes at the individual level should be based on individual-level data; conclusions about group-level processes should be based on data collected about groups. When this rule is violated, you can be misled about the existence of an association between two variables.

A researcher who draws conclusions about individual-level processes from group-level data could be making what is termed an **ecological fallacy** (see Table 6.1). The conclusions may or may not be correct, but we must recognize the fact that group-level data do not necessarily reflect solely individual-level processes. For example, a researcher may examine factory records and find that the higher the percentage of unskilled workers in factories, the higher the rate of employee sabotage in those factories. But the researcher would commit an ecological fallacy if they then concluded that individual unskilled factory workers are more likely to engage in sabotage. This conclusion is about an individual-level causal process (the relationship between the occupation and criminal propensities of individuals), even though the data describe groups (factories). It could be that white-collar workers are the ones more likely to commit sabotage in factories with more unskilled workers, perhaps because the white-collar workers feel they won't be suspected in these settings.

TABLE 6.1 ■ Errors in Causal Conclusions

		You make conclusions about	
		Groups	**Individuals**
You collect data from	Groups	More homogeneous groups tend to have stronger social bonds.	Groups with a higher average age are more conservative, so older people are more conservative. *Possible Ecological Fallacy*
	Individuals	Students who socialize more have lower grades, so schools with more social engagement will have poorer student performance. *Possible Reductionist Fallacy*	Older people tend to be more conservative.

Conversely, when data about individuals are used to make inferences about group-level processes, a problem occurs that can be thought of as the mirror image of the ecological fallacy: the **reductionist fallacy**, also known as *reductionism*, or the *individualist fallacy* (see Table 6.1). For example, a reductionist explanation of individual violence would focus on biological factors, such as genes or hormones, rather than on the community's level of social control. Similarly, a reductionist explanation of behavior problems in grade school classrooms would focus on the children's personalities rather than on classroom structure, teacher behavior, or the surrounding neighborhood.

There is, of course, nothing inherently wrong with considering biological, psychological, or other factors that affect individuals when explaining behavior in social contexts (Sayer 2003).

The key issue is whether there are properties of groups or other aggregates that are more than "the sum of their parts." If there are such "higher level properties," we can speak of "**emergence**" at the group level, and we need to be careful to avoid a reductionist *fallacy*.

The fact that errors in causal reasoning can be made in this way should not deter you from conducting research with group-level data nor make you unduly critical of researchers who make inferences about individuals based on group-level data. When considered broadly, many research questions point to relationships that could be manifested in many ways and on many levels. Robert J. Sampson et al. (2018) emphasize structural disadvantage as a key predictor of variation in violence, with state-level, city-level and community-level aspects of structural disadvantage being important. They also caution that "we do not deny the importance of individual-level analyses of race and crime or claim that *only* community factors explain crime" (p. 29), and they urge attention to policies that compensate for disadvantages experienced by poor individuals, such as providing cash assistance and job opportunities (pp. 27–28).

The solution is to know what the units of analysis and units of observation were in a study and to consider these in weighing the credibility of the researcher's conclusions. The goal is not to reject out of hand conclusions that refer to a level of analysis different from what was actually studied. Instead, the goal is to consider the likelihood that an ecological fallacy or a reductionist fallacy has been made when estimating the causal validity of the conclusions.

RESEARCH IN THE NEWS
WHY DOES LOUISIANA LEAD THE UNITED STATES IN MURDERS?

Louisiana has reported the highest murder rate in the United States for 31 years. History suggests that part of the explanation could be the enduring influence of the Southern "frontier justice" tradition of people taking the law into their own hands and a regional "honor justice" mentality that is quick to take offense at any sign of disrespect. Former New Orleans mayor Marc Morial (1994 to 2002) explained the problem as "poverty and its twin sister or brother of mass incarceration . . . and it's easy access to guns." But "criminologists tread carefully in inferring causation" and have examined other factors unique to Louisiana. Slavery on the Louisiana sugar plantations was much more barbaric compared to cotton plantations in other states, fueling a culture of violence that continued with racial massacres and lynching after the Civil War. Louisiana has also been termed "the world's prison capital" for its high incarceration rate, motivated in part by financial incentives to local sheriffs for keeping their prisons full. The violent environment of prisons in turn increases the likelihood of further violent acts after release from prison. And ATF (Bureau of Alcohol, Tobacco, Firearms and Explosives) figures suggest a high rate of illegal or stolen weapons in Louisiana.

For Further Thought?

1. What steps could you take as a researcher to distinguish the effects on crime of prior incarceration, poverty, and education? What other influences do you think should be examined?
2. How much weight do you think should be given to historical factors in explanations of current social patterns? Have you seen evidence of present effects of cultural patterns in the past?

News source: Asher, Jeff, Ben Horwitz, and Toni Monkovic. 2021. "Why Does Louisiana Lead the United States in Murders?" *The New York Times*, February 16:A16.

CROSS-SECTIONAL AND LONGITUDINAL DESIGNS

Do you want to describe or understand social phenomena in the present? If you want to assess support for a candidate a month before a local election, you need to collect your data at that time. If you want to describe the level of violence in a sample of communities, your focus might be on the current year. If the focus of your investigation is on the present or some other specific limited period, your research design will be cross-sectional. In **cross-sectional research designs**, all data are collected at one point in time. However, if you want to track changes in support for a candidate during the entire campaign period or describe variation during the past decade in the level of violence in a sample of communities, you will need to collect data from different times across the entire period you are investigating. In **longitudinal research designs**, data are collected at two or more points in time.

Therefore, the research question determines whether a researcher needs to collect cross-sectional or longitudinal data. If the research question concerns only the here and now, there is no need for longitudinal data. However, it is also important to recognize that any research question involving a causal analysis—about what causes what—creates an issue about change over time. Identifying the **time order** of effects—what happened first and so on—is critical for developing a causal analysis but can be an insurmountable problem with a cross-sectional design. In longitudinal research designs, identification of the time order of effects can be quite straightforward.

Cross-Sectional Designs

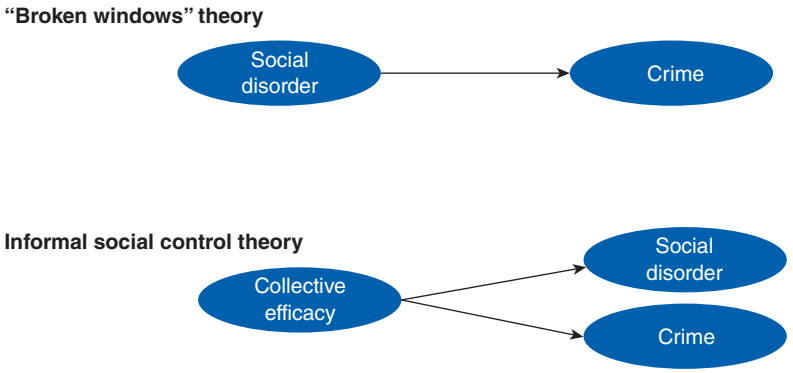
Much of the research you have encountered so far in this text—the observations about computer use in Chapter 1, the surveys about the meaning of tolerance in Chapter 4 and of homeless persons in Chapter 5—has been cross-sectional. Although each of these studies took some time to carry out, researchers measured the actions, attitudes, and characteristics of respondents at only one point in time.

Sampson and Raudenbush (1999) used an ambitious cross-sectional design to investigate the effect of visible public social and physical disorder on the crime rate in Chicago neighborhoods. Their theoretical framework focused on the concept of informal social control—the ability of residents to regulate social activity in their neighborhoods through their collective efforts according to desired principles. The researchers believed that informal social control would vary between neighborhoods, and they hypothesized that the strength of informal social control rather than just the visible signs of disorder would explain variation in crime rates. Sampson and Raudenbush contrasted this prediction with the "broken windows" theory: the belief that signs of disorder themselves cause crime. In the theory Sampson and Raudenbush proposed, both visible disorder and crime were consequences of low levels of informal social control (measured with an index of collective efficacy). One did not cause the other (see Figure 6.4).

Sampson and Raudenbush (1999) measured visible disorder through direct observation: Trained observers rode slowly around every street in 196 Chicago census tracts. Sampson and Raudenbush also conducted a survey of residents and examined police records. Both survey responses and police records were used to measure crime levels. The level of neighborhood informal social control and other variables were measured with the average resident responses to several survey questions. Both the crime rate and the level of social and physical disorder varied between neighborhoods in relation to the level of informal social control. Informal social control (collective efficacy) was a much more important factor in the neighborhood crime rate than was visible social and physical disorder, measured at the same time (see Figure 6.5).

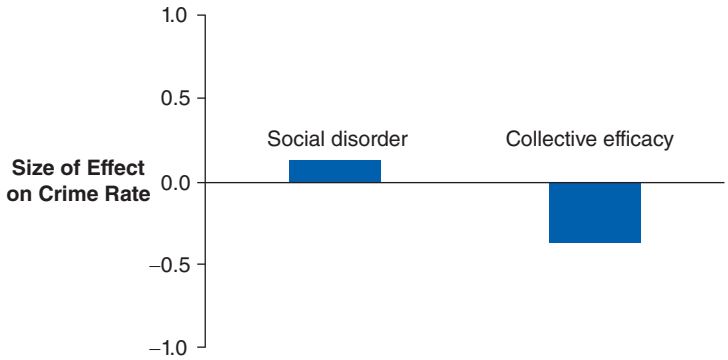
But note that we are left with a difficult question about the relations Sampson and Raudenbush identified between the variables they measured. Did neighborhoods that developed

FIGURE 6.4 ■ The Effect of Informal Social Control



Source: Based on Sampson and Raudenbush (1999).

FIGURE 6.5 ■ Effect of Social Disorder and Collective Efficacy on Personal Violent Crime Rates



Source: Based on Sampson and Raudenbush (1999).

higher levels of informal social control then experience reductions in crime, or was it the other way around: Did neighborhoods that experienced a drop in crime then develop higher levels of informal social control? After all, if you are afraid to leave your apartment because you fear crime, you can't do a very good job of keeping an eye on things or attending community meetings. Maybe the crime reduction made residents feel safe to engage in more informal social control efforts rather than just calling the police. Because of uncertainty like this, it is almost always better to use a longitudinal research design rather than a cross-sectional research design to answer questions about causal effects.

There are four special circumstances in which we can be more confident in drawing conclusions about time order—and hence conclusions about causality—on the basis of cross-sectional data. Because in these special circumstances the data can be ordered in time, they might even be thought of as longitudinal designs (Campbell 1992). These four special circumstances are as follows:

1. *The independent variable is fixed at some point before the variation in the dependent variable.* So-called demographic variables that are determined at birth—such as sex, race, and age—are fixed in this way. So are variables such as education and marital

status, if we know when the value of cases on these variables was established. If we also know that the value of cases on the dependent variable was set some time afterward, the time order of variation in the independent and dependent variables is clear. For example, say we hypothesize that education influences the type of job individuals have. If we know that respondents completed their education before taking their current jobs, we will satisfy the time-order requirement, even if we were to measure education at the same time we measure the type of job. However, if some respondents possibly went back to school as a benefit of their current jobs, the time-order t would not be clear.

2. *We believe that respondents can give us reliable reports of what happened to them or what they thought at some earlier point in time.* Julie Horney et al. (1995) provide an interesting example of the use of such retrospective data. The researchers wanted to identify how criminal activity varies in response to changes in life circumstances. They interviewed 658 newly convicted male offenders sentenced to a Nebraska state prison. In a 45- to 90-minute interview, Horney et al. recorded each incarcerated person's report of their life circumstances and of their criminal activities for the preceding two to three years. The researchers then found that criminal involvement appeared to be preceded by adverse changes in life circumstances, such as marital separation or drug use. Retrospective data are often inadequate for measuring variation in past feelings, events, or behaviors, however, because we may have difficulty recalling what we have felt or what has happened in the past and what we do recall is likely to be influenced by what we feel in the present (Elliott, Holland, and Thomson 2008:229). For example, retrospective reports by both adults suffering from alcoholism and their parents appear to overestimate greatly the frequency of childhood problems (Vaillant 1995). People cannot report reliably the frequency and timing of many past events, from hospitalization to hours worked. However, retrospective data tend to be reliable when they concern major, persistent experiences in the past, such as what type of school someone went to or how a person's family was structured (Campbell 1992).

3. *Our measures are based on the records that contain information on cases in earlier periods.* Government, agency, and organizational records can be excellent sources of time-ordered data after the fact. However, sloppy record keeping and changes in data collection policies can lead to inconsistencies, which must be considered. Another weakness of such archival data is that they usually contain measures of only a fraction of the variables we think are important. And of course we must have access to the data in a way that preserves the confidentiality of information about individuals.

4. *We know that the value of the dependent variable was similar for all cases before the treatment.* For example, we may hypothesize that a training program (independent variable) improves the English-speaking abilities (dependent variable) of a group of recent immigrants. If we know none of the immigrants could speak English before enrolling in the training program, we can be confident that any subsequent variation in their ability to speak English did not precede exposure to the training program. This is one way that traditional experiments establish time order: Two or more equivalent groups are formed before exposing one of them to some treatment.

Longitudinal Designs

In longitudinal research, data are collected that can be ordered in time. By measuring the value of cases on an independent variable and a dependent variable at different times, the researcher

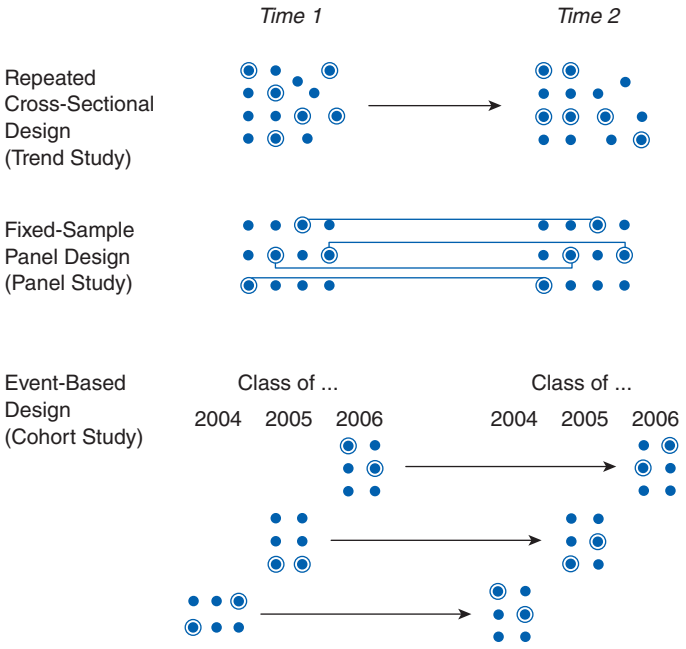
can determine whether variation in the independent variable precedes variation in the dependent variable.

In some longitudinal designs, the same sample (or panel) is followed over time; in other designs, sample members are rotated or completely replaced. The population from which the sample is selected may be defined broadly, as when a longitudinal survey of the general population is conducted. Or the population may be defined narrowly, as when the members of a specific age group are sampled at multiple points in time. The frequency of follow-up measurement can vary, ranging from a before-and-after design with just the one follow-up to studies in which various indicators are measured every month for many years.

Certainly, it is more difficult to collect data at two or more points in time than at one time. Quite frequently researchers simply cannot or are unwilling to delay completion of a study for even one year to collect follow-up data. But think of the many research questions that really should involve a much longer follow-up period: What is the impact of job training on subsequent employment? How effective is a school-based program in improving parenting skills? How has the unemployment rate changed as monetary policy has changed? Under what conditions do traumatic experiences in childhood result in mental illness? It is safe to say that we will never have enough longitudinal data to answer many important research questions. Nonetheless, the value of longitudinal data is so great that every effort should be made to develop longitudinal research designs when they are appropriate for the research question asked. The following discussion of the three major types of longitudinal designs will give you a sense of the possibilities (see Figure 6.6).

Repeated cross-sectional designs (trend studies). Studies that use a **repeated cross-sectional design**, also known as *trend studies*, have become fixtures of the political arena around election time. Particularly in presidential election years, we have all become accustomed to reading weekly, even daily, reports on the percentage of the population that supports each candidate. Similar polls are conducted to track sentiment on many other social issues. For example, a 1993

FIGURE 6.6 ■ Three Types of Longitudinal Design



poll reported that 52% of adult Americans supported a ban on the possession of handguns, compared with 41% in a similar poll conducted in 1991. According to pollster Louis Harris, this increase indicated a "sea change" in public attitudes (cited in Barringer 1993). Another researcher said, "It shows that people are responding to their experience [of an increase in handgun-related killings]" (cited in Barringer 1993:A14).

Repeated cross-sectional surveys are conducted as follows:

1. A sample is drawn from a population at Time 1, and data are collected from the sample.
2. As time passes, some people leave the population and others enter it.
3. At Time 2, a different sample is drawn from this population.

These features make the repeated cross-sectional design appropriate when the goal is to determine whether a population has changed over time. Matthew Desmond, Andrew V. Papachristos, and David S. Kirk (2016) used this design to study change over time in 911 calls (see Chapter 2). Has racial tolerance increased among Americans in the past 20 years? Are employers more likely to pay maternity benefits today than they were in the 1950s? These questions also concern change in the population, not just change in individuals within the population. We may want to know whether racial tolerance increased in society, not whether this change was due to migration that brought more racially tolerant people into the country or to individual U.S. citizens becoming more tolerant. We may ask whether employers overall are more likely to pay maternity benefits today than they were yesterday, not whether any such increase was due to recalcitrant employers going out of business or to individual employers changing their maternity benefits. When we do need to know whether individuals in the population changed, we must turn to a panel design.

Fixed-sample panel designs (panel studies) allow us to identify changes in individuals, groups, or whatever we are studying. This is the process for conducting **fixed-sample panel designs**:

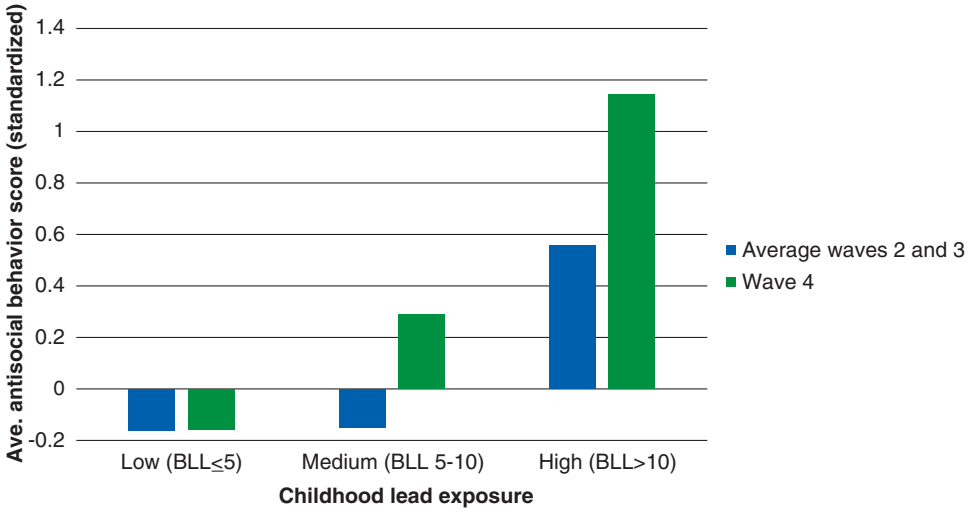
1. A sample (called a panel) is drawn from a population at Time 1, and data are collected from the sample.
2. As time passes, some panel members become unavailable for follow-up, and the population changes.
3. At Time 2, data are collected from the same people as at Time 1 (the panel)—except for those people who cannot be located.

Because a panel design follows the same individuals, it is better than a repeated cross-sectional design for testing causal hypotheses.

For example, Robert J. Sampson and Alix X. Winter (2018) used a fixed-sample panel design to investigate the effect of childhood lead exposure on subsequent adolescent delinquency. They studied children born in Chicago households through interviews with their caregivers in the mid-1990s and then several additional times, including when the children had become adolescents (16–18 years old). Sampson and Winter (2018:292) found that children who had been exposed to lead as children (as measured in a blood test) were more likely to engage in antisocial behavior while still children (waves 2 and 3) and as older adolescents (wave

4) (see Figure 6.7). The researchers' use of a panel design allowed them to be confident that the lead exposure occurred before the delinquent behavior that it was hypothesized to cause.

FIGURE 6.7 ■ Childhood Blood Lead Level (prior to age 6) and Adolescent Antisocial Behavior



Source: Sampson, Robert J. and Delaram Takyar. 2018. "Poisoned Development: Assessing Childhood Lead Exposure as a Cause of Crime in a Birth Cohort Followed Through Adolescence." *Criminology*, 56(2):280.

Despite their value in establishing time order of effects, panel studies are a challenge to implement successfully, so they often are not even attempted. There are two major difficulties:

1. *Expense and attrition.* It can be difficult and very expensive to keep track of individuals over a long period, and inevitably the proportion of panel members who can be located for follow-up will decline over time. Panel studies often lose more than one quarter of their members through attrition (Miller and Salkind 2002:321), and those who are lost were often not necessarily like those who remain in the panel. As a result, a high rate of subject attrition may mean the follow-up sample will no longer be representative of the population from which it was drawn and may no longer provide a sound basis for estimating change. Subjects who were lost to follow-up may have been those who changed the most or the least over time. For example, between 5% and 66% of subjects are lost in substance abuse prevention studies, and the dropouts typically had begun the study with higher rates of tobacco and marijuana use (Snow, Tebes, and Arthur 1992:804).

   It does help to compare the baseline characteristics of those who are interviewed at follow-up with characteristics of those lost to follow-up. If these two groups of panel members were not very different at baseline, it is less likely that changes had anything to do with characteristics of the missing panel members. Even better, subject attrition can be reduced substantially if sufficient staff can be used to keep track of panel members. In their panel study, Sampson and Winter (2018) were able

to interview 67% of caregivers from their initial sample more than a decade after the first interviews.

2. *Subject fatigue.* Panel members may grow weary of repeated interviews and drop out of the study, or they may become so used to answering the standard questions in the survey that they start giving stock answers rather than thinking about their current feelings or actions (Campbell 1992). This problem is called **subject fatigue**. Fortunately, subjects do not often seem to become fatigued in this way, particularly if the research staff have maintained positive relations with the subjects. For example, at the end of an 18-monthlong experimental study of housing alternatives for persons diagnosed with mental illness who had been homeless, only three or four individuals (out of 93 who could still be located) refused to participate in the fourth and final round of interviews. The interviews took a total of about five hours to complete, and participants received about $50 for their time (Schutt, Goldfinger, and Penk 1997).

   Because panel studies are so useful, social researchers have developed increasingly effective techniques for keeping track of individuals and overcoming subject fatigue. But when resources do not permit use of these techniques to maintain an adequate panel, repeated cross-sectional designs usually can be employed at a cost that is not a great deal higher than that of a one-time-only cross-sectional study. The payoff in explanatory power should be well worth the cost.

Event-based designs (cohort studies). In an **event-based design**, often called a *cohort study*, the follow-up samples (at one or more times) are selected from the same **cohort**—people who all have experienced a similar event or a common starting point. Examples include the following:

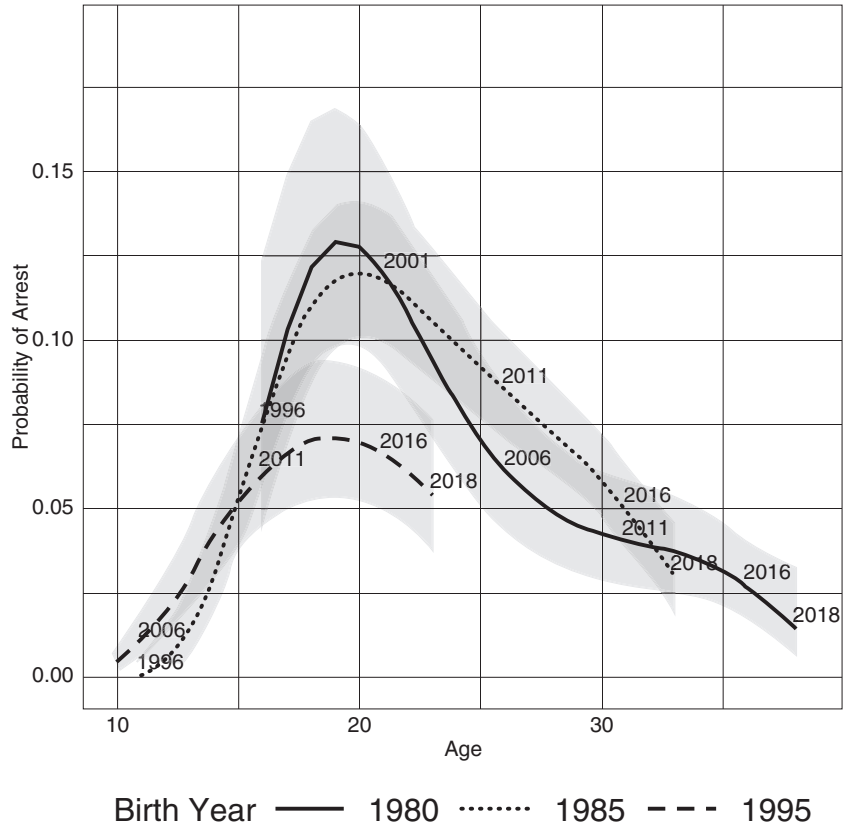
- *Birth cohorts*—those who share a common period of birth (those born in the 1970s, 1980s, 1990s, etc.)
- *Seniority cohorts*—those who have worked at the same place for about 5 years, about 10 years, and so on
- *School cohorts*—freshmen, sophomores, juniors, and seniors

An event-based design can be a type of repeated cross-sectional design or a type of panel design. In an event-based repeated cross-sectional design, separate samples are drawn from the same cohort at two or more different times. In an event-based panel design, the same individuals from the same cohort are studied at two or more different times. Comparing findings between different cohorts can help reveal the importance of the social or cultural context that the different cohorts experienced (Elliott, Holland, and Thomson 2008:230).

Event-based research can improve the identification of causal effects compared with cross-sectional designs. Roland Neil and Robert J. Sampson (2021) used an event-based (cohort) design to identify the effect of changing social circumstances on rates of arrest. Using households recruited for the Project on Human Development in Chicago Neighborhoods (PHDCN), Neil and Sampson studied the arrest records of four cohorts of individuals born between 1979 and 1996, after these cohorts grew up in periods of changing law enforcement practices and falling crime levels.

Among those in the PHDCN sample who remained in Chicago, the probability of arrest at different ages was markedly lower for the cohorts born in 1996 or later. Neil and Sampson (2021) then show that this difference is explained by both the reduced level of crime and less intense patterns of law enforcement experienced by these more recent cohorts (Figure 6.8).

FIGURE 6.8 ■ Probability of Arrest at Different Ages for Three Birth Cohorts



Source: Roland Neil and Robert J. Sampson. 2021. "The Birth Lottery of History: Arrest Over the Life Course of Multiple Cohorts Coming of Age, 1995–2018." *American Journal of Sociology,* 5:1153.

NOMOTHETIC OR IDIOGRAPHIC CAUSAL EXPLANATIONS

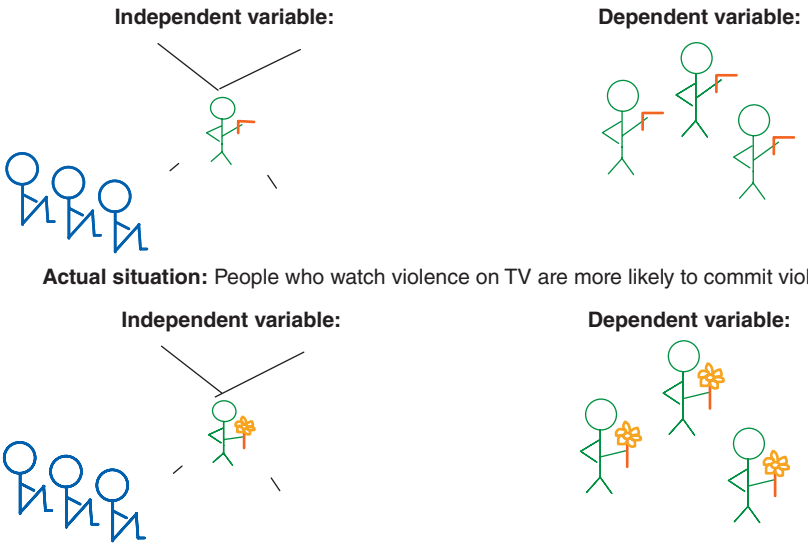
A cause is an explanation for some characteristic, attitude, or behavior of groups, individuals, or other entities (such as families, organizations, or cities) or for events. Most social scientists seek causal explanations that reflect tests of the types of hypotheses with which you are familiar (see Chapter 2): The independent variable is the presumed cause, and the dependent variable is the potential effect. For example, the study by Sampson and Winter (2018) tested whether childhood lead exposure (the independent variable) leads to delinquent behavior (the dependent variable). (As you know, they concluded that it did, although it did not also result in more arrests.) This type of causal explanation is termed *nomothetic*.

A different type of cause is the focus of some qualitative research (Chapter 10), some historical and comparative research (Chapter 15), and our everyday conversations about causes. In this type of causal explanation, termed *idiographic*, individual events or the behaviors of individuals are explained with a series of related, prior events. For example, you might explain a particular crime as resulting from several incidents in the life of the perpetrator that resulted in a tendency toward violence, coupled with stress resulting from a failed marriage and maybe also due to a chance meeting.

Nomothetic Causal Explanations

A **nomothetic causal explanation** is one involving the belief that variation in an independent variable will be followed by variation in the dependent variable when all other things are equal (**ceteris paribus**). Researchers who claim a **causal effect** from a **nomothetic perspective** have concluded that the value of cases on the dependent variable differs from what their value would have been in the absence of variation in the independent variable. For instance, researchers might claim the likelihood of committing violent crimes is higher for individuals who were abused as children than it would be if these same individuals had not been abused as children. Or researchers might claim the likelihood of committing violent crimes is higher for individuals exposed to media violence than it would be if these same individuals had not been exposed to media violence. The situation as it would have been in the absence of variation in the independent variable is termed the **counterfactual** (see Figure 6.9).

FIGURE 6.9 ■ The Counterfactual in Causal Research



Actual situation: People who watch violence on TV are more likely to commit violent acts.

Counterfactual situation: The same people watch nonviolent TV shows at the same time, in the same circumstances. They are not more likely to commit violent acts.

Example of a Nomothetic Causal Effect Individuals arrested for domestic assault tend to commit fewer subsequent assaults than do similar individuals who are accused in the same circumstances but not arrested.

Of course, the fundamental difficulty with this perspective is that we never really know what would have happened at the same time to the same people (or groups, cities, etc.) if the independent variable had not varied—because it did (Shrout 2011:4–5). We can't rerun real-life scenarios (King, Keohane, and Verba 1994). We could observe the aggressiveness of people's behavior before and after they were exposed to media violence. But this comparison involves an earlier time, when, by definition, the people and their circumstances were not exactly the same.

But we do not need to give up hope! Far from it. We can design research to create conditions that are comparable indeed, so that we can confidently assert our conclusions ceteris paribus—other things being equal. We can examine the impact on the dependent variable of variation

in the independent variable alone, even though we will not be able to compare the same people at the same time in exactly the same circumstances, except for the variation in the independent variable. And by knowing the ideal standard of comparability (the counterfactual), we can improve our research designs and strengthen our causal conclusions, even when we cannot come so close to living up to the meaning of ceteris paribus.

Actual situation: People who watch violence on TV are more likely to commit violent acts.

Counterfactual situation: The same people watch nonviolent TV shows at the same time, in the same circumstances. They are not more likely to commit violent acts.

Quantitative researchers seek to test nomothetic causal explanations with either experimental or nonexperimental research designs. However, the way in which experimental and nonexperimental designs attempt to identify causes differs quite a bit. It is very hard to meet some of the criteria for achieving valid nomothetic causal explanations using a nonexperimental design. Most of the rest of this chapter is devoted to a review of these causal criteria and a discussion of how experimental and nonexperimental designs can help establish them.

Qualitative researchers can also seek nomothetic causal explanations. Bas Teirolf, Edith Geurts, and Majone Steketee (2021) collected both quantitative and qualitative data to explain the effect of the pandemic on domestic violence in families being seen for child protective services in the Netherlands. Comparing questionnaire responses from 159 parents and children just before the pandemic and 87 recruited after the pandemic lockdown began, Tierolf, Geurts, and Steketee found no quantitative difference in the level of domestic violence. Qualitative interviews with 39 of the family respondents and 13 care professionals did indicate higher levels of stress during the pandemic due to problems like lack of care for children and combining working at home with childcare but also identified sources of resilience in parental adaptations.

Idiographic Causal Explanations

The other meaning of the term *cause* is one that we have in mind very often in everyday speech. This is **idiographic causal explanation**: the concrete, individual sequence of events, thoughts, or actions that resulted in a particular outcome for a particular individual or that led to a particular event (Hage and Meeker 1988). An idiographic explanation also may be termed an *individualist* or a *historicist* explanation. It is more likely to be used for causal explanations in qualitative research.

A **causal effect** from an **idiographic perspective** includes statements of initial conditions and then relates a series of events at different times that led to the outcome, or causal effect. This narrative or story is the critical element in an idiographic explanation, which may therefore be classified as narrative reasoning (Richardson, 1995:200–201). Idiographic explanations focus on particular social actors, in particular social places, at particular social times (Abbott 1992). Idiographic explanations are also typically very concerned with context—with understanding the particular outcome as part of a larger set of interrelated circumstances. Idiographic explanations can thus be termed *holistic*. The following idiographic explanation of the history of one individual's homeless history by Hirschi (1989) takes into accunt multiple circumstances, including childhood experiences of neglect and of some support.

Example of an Idiographic Causal Effect An individual is neglected by her parents but has a supportive grandparent. She comes to distrust others, has trouble in school, is unable to keep a job, and eventually becomes homeless. She subsequently develops a supportive relationship with a shelter case manager, who helps her find a job and regain her housing (based on Hirsch 1989).

Elijah Anderson's (1990) field research in a poor urban community produced a narrative, idiographic account of how drug addiction can result in a downward slide into residential instability and crime:

> When addicts deplete their resources, they may go to those closest to them, drawing them into their schemes. . . . The family may put up with the person for a while. They provide money if they can. . . . They come to realize that the person is on drugs. . . . Slowly the reality sets in more and more completely, and the family becomes drained of both financial and emotional resources. . . . Close relatives lose faith and begin to see the person as untrustworthy and weak. Eventually the addict begins to "mess up" in a variety of ways, taking furniture from the house [and] anything of value. . . . Relatives and friends begin to see the person . . . as "out there" in the streets. . . . One deviant act leads to another. (pp. 86–87)

An idiographic explanation can also be developed from the narration of an individual. For example, Carole Cain interviewed AA (Alcoholics Anonymous) members about their experiences to learn how they construct their identities as alcoholics. In one interview, excerpted by Catherine Kohler Riessman (2008:71), "Hank" describes some of his experiences with drinking:

> One morning he found he could not get up even after several drinks. … When he did get up, he found AA, although he cannot remember how he knew where to go…. From the morning when he contacted AA, he did not drink again for over five years. … Life improved, he got himself in better shape and got back together with his wife. After several years, the marriage broke up again, and in anger with his wife, he went back to drinking for another five years.

An idiographic explanation such as Anderson's or Cain's pays close attention to time order and causal mechanisms. Nonetheless, it is difficult to make a convincing case that one causal narrative should be chosen over an alternative narrative (Abbott 1992). Does low self-esteem result in vulnerability to the appeals of drug dealers, or does a chance drug encounter precipitate a slide in self-esteem? Did drinking lead Hank to go to AA, or did he start drinking more because he knew it would push him to go to AA? Did his drinking start again because his marriage broke up, or did his orientation lead his wife to leave and to his renewed drinking? The prudent causal analyst remains open to alternative explanations.

Idiographic explanation is deterministic, focusing on what caused a particular event to occur or what caused a particular case to change. As in nomothetic explanations, idiographic causal explanations can involve counterfactuals, by trying to identify what would have happened if a different circumstance had occurred. But unlike in nomothetic explanations, in idiographic explanations, the notion of a probabilistic relationship, an average effect, does not really apply. A deterministic cause is believed to have an effect in the specific case under consideration.

CRITERIA FOR NOMOTHETIC CAUSAL EXPLANATIONS

How the research is designed influences our ability to draw causal conclusions. In this section, I introduce the features that need to be considered in a research design to evaluate how well it can support nomothetic causal conclusions.

Three criteria must be considered when deciding whether a causal connection exists. When a research design leaves one or more of the criteria unmet, we may have some important doubts about causal assertions that the researcher may have made. These three criteria are generally

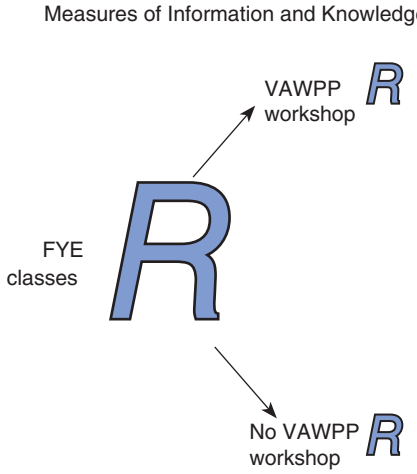
considered the most important bases for identifying a nomothetic causal effect: (1) empirical association, (2) appropriate time order, and (3) nonspuriousness (Hammersley 2008:43). The features of experimental research designs are particularly well suited to meeting these criteria and for testing nomothetic causal explanations. However, we must also consider the degree to which these criteria are met when evaluating nonexperimental research that is designed to test causal hypotheses.

Two other issues that I introduce as "cautions" in this chapter are a bit different. They are not necessary for establishing that a causal connection exists, but they help us understand it better. If we have not identified a causal mechanism, the first caution, we do not understand fully *why* a causal connection exists. The second caution is to specify the context in which a causal effect occurs because by understanding *when* or *under what conditions* the causal effect occurs, we will understand better what that effect *is.* Providing information about both mechanism and context can considerably strengthen causal explanations (Hammersley 2008:44–45).

In the following subsections, I will indicate how researchers attempt to meet the three criteria and address the two cautions with research projects that illustrate both experimental and nonexperimental designs. Illustrations of experimental design features will rely on Lauren Donais, Bill Simonsen, and Nicole Simonsen's (2019) evaluation of the effect of gender-based violence prevention workshops. Most illustrations of nonexperimental design features will refer to the study by Sampson and Winter (2018) of the effect of childhood lead exposure on crime, which I introduced earlier.

Donais and her coauthors (2019) delivered a Violence Against Women Prevention Program (VAWPP) workshop in the University of Connecticut's small first-year experience (FYE) classes in the Fall of 2015. The classes were divided randomly into two groups for the experiment. Students (912) in one randomly selected half of the classes were given a questionnaire to complete about the information included in the VAWPP workshop before their class participated in the workshop, while students (1045) in the other half of the classes completed the questionnaire after they participated in the VAWPP workshop (see Figure 6.10).

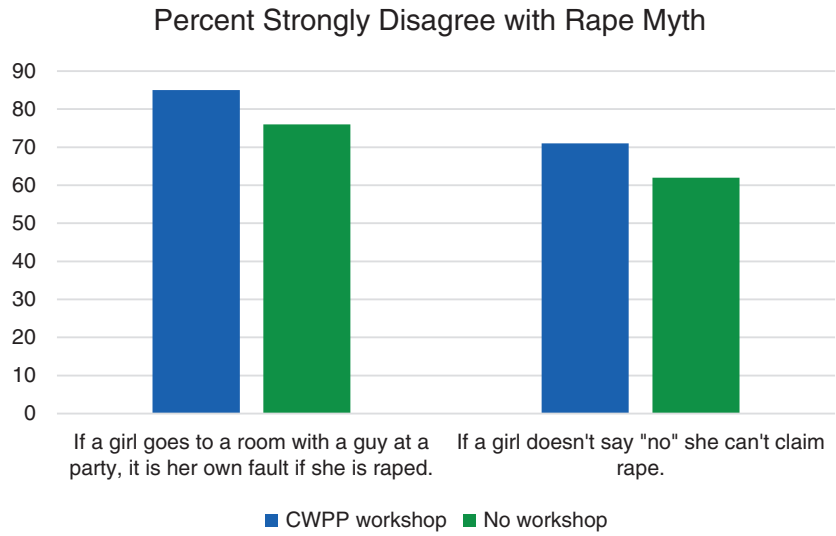
FIGURE 6.10 ■ Experiment to Test the Effect of Gender-Based Violence Prevention Workshops



Source: Adapted from Donais, Lauren, Bill Simonsen, and Nicole Simonsen. 2019. "Gender-Based Violence Prevention Workshops: An Experimental Evaluation of Efficacy." *International Journal of Public Administration* 42:840–854.

The students who had experienced the VWPP workshop in an FEW class were less likely to agree with "rape myth" statements that put the responsibility for a rape on the woman. Figure 6.11 shows the increase in disagreement with two of the "rape myth" statements among workshop participants. There were similar differences in response to other such statements and to questions about knowledge and confidence concerning appropriate behavior in situations requiring consent. Donais, Simonsen, and Simonsen (2019) concluded that the workshops reduced rape myth acceptance and related knowledge and attitudes.

FIGURE 6.11 ■ Percent Strongly Disagree with Rape Myth



Source: Adapted from Donais, Lauren, Bill Simonsen, and Nicole Simonsen. 2019. "Gender-Based Violence Prevention Workshops: An Experimental Evaluation of Efficacy." *International Journal of Public Administration* 42:846.

Was this causal conclusion justified? How confident can we be in its internal validity? Do you know if the students who participated in the CWPP workshop before answering the rape myth questions were the same before the workshop as students who answered the rape myth questions before they participated in the workshop? What about external validity? Do you think that college students' verbal reactions after a workshop in a college class can be generalized to their behavior in actual encounters? In the next sections, I will show how well the features of the research designs used by Donais and her coauthors meet the criteria for nomothetic causal explanation and thus determine the confidence we can place in their causal conclusions. I will also identify those features of a *true experiment* that make this research design particularly well suited to testing nomothetic causal hypotheses.

Association

We say there was an **association** between workshop participation and rejection of rape myths in Donais et al.'s (2019) experiment because the level of agreement with the "rape myth" statements varied according to whether students had participated in the workshop or not before they completed the questionnaire. An empirical (or observed) association between the independent and dependent variables is the first criterion for identifying a nomothetic causal effect.

We can determine whether an association exists between the independent and dependent variables in a true experiment because there are two or more groups that differ in their value on the independent variable. One group receives some "treatment"—such as reading a message—that manipulates the value of the independent variable. This group is termed the *experimental group.* In a simple experiment, there may be one other group that does not receive the treatment; it is termed the *control group.* In the Donais study, the students who had participated in the workshop were the experimental group, while the students who had not yet participated in the workshop were the control group. More complex experiments may compare more groups that represent multiple values of the independent variable or combinations of the values of more than two independent variables.

In nonexperimental research, the test for an association between the independent and dependent variables is like that used in experimental research—seeing whether values of cases that differ on the independent variable tend to differ in the dependent variable. The difference with nonexperimental research designs is that the independent variable is not a treatment to which the researcher assigns some individuals. In their nonexperimental study of the effect of childhood lead exposure, Sampson and Winter (2018) compared the average number of delinquent acts and the number of arrests—the dependent variables—for adolescents who had been exposed to lead as children (as indicated by childhood blood tests)—the independent variable—with the average number of delinquent acts and number of arrests by adolescents who had not been exposed to lead as children, but this was not an experimental study and so the researchers had not assigned children to be exposed to lead or not (of course, it would be unethical for any researcher to do so).

But this is the time to be forewarned so that you can be forearmed: Have you heard the adage "correlation does not prove causation"? It highlights the fact that an association between two variables can occur for many reasons other than a causal effect of one variable on the other variable, including simple coincidence. Don't fall into that easy trap. The home team was winning more games and housing prices were rising? *So what? Correlation does not prove causation!*

Time Order

Association is a necessary criterion for establishing a causal effect, but it is not sufficient. We must also ensure the variation in the dependent variable occurred after the variation in the independent variable. This is the criterion of time order. Our research design determines our ability to determine time order.

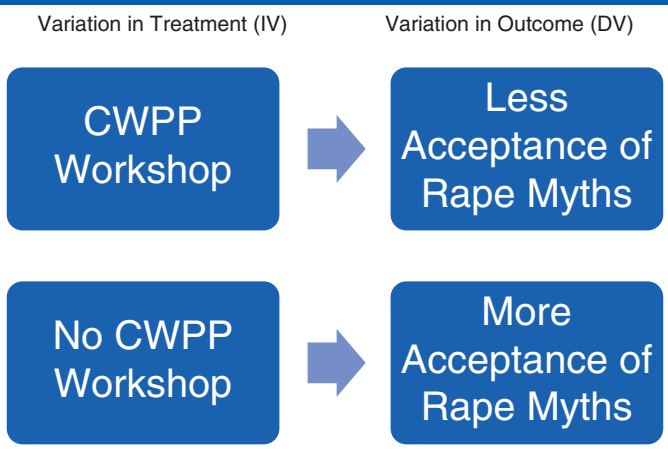
Experimental Designs

In a true experiment like that conducted by Donais et al. (2019), the researcher determines the time order. Students in their treatment group were in the random half of the FYE classes that hosted the CWPP workshop before the students completed the questionnaire about acceptance of rape myths and the other dependent variables, while students in the control condition were in the random half of the FYE classes that had not yet had the CWPP workshop when they completed the same questionnaire. Therefore, we know the variation in the independent variable (participation in a CWPP workshop) occurred before the measurement of the hypothesized outcome, the dependent variables. This is how time order is established by Donais et al.'s research design (Figure 6.12).

Nonexperimental Designs

You have already learned that nonexperimental research designs can be either cross-sectional or longitudinal. In most cross-sectional studies, conclusions about causation must be more tentative

FIGURE 6.12 ■ Time Order in Experimental Design



because the time order of effects is uncertain. Imagine conducting a cross-sectional survey of students in which you include the same questions as in the Donais et al. (2019) study about acceptance of rape myths and a question asking about prior participation in a CWPP workshop. If we find an association between rejection of rape myths and participation in a CWPP workshop, we won't know if students who were more likely to reject rape myths were more likely to participate in a CWPP workshop than students who were less likely to reject rape myths. In other words, we won't know if variation in the hypothesized independent variable, participation in a CWPP workshop, occurred after variation in the hypothesized dependent variable, acceptance of rape myths. The association we found in this nonexperimental cross-sectional design does not give us very strong evidence that the workshop participation caused the variation in acceptance of rape myths.

Longitudinal designs that do not involve an experiment can still establish the time order of effects and so provide stronger evidence about causality. For example, Sampson and Winter (2018) measured lead exposure of children and then measured the delinquent behaviors of these same children during their adolescent years. Their panel study design allowed confidence in the conclusion that the variation in the independent variable—lead exposure—occurred before the delinquent behavior that it was hypothesized to cause. As you learned in the earlier section in this chapter on cross-sectional designs, time order can still be established in some cross-sectional designs. If Sampson and Winter had only used a cross-sectional study of adolescents but also had obtained medical records that had indicated levels of lead in their blood when they were children, they could still have been confident that the lead exposure occurred before the delinquent behavior (see earlier section on cross-sectional designs in this chapter).

Nonspuriousness

Nonspuriousness is another essential criterion for establishing the existence of a causal effect of an independent variable on a dependent variable; in some respects, it is the most important criterion. We say that a relationship between two variables is not spurious when it is not caused by variation in a third variable.

Do storks bring babies? If you mistakenly believe that correlation proves causation, then you might think so. The more storks that appear in certain districts in Holland, the more babies are born. But the association in Holland between number of storks and number of babies is spurious.

In fact, both the number of storks and the birthrate are higher in rural districts than in urban districts. The rural or urban character of the districts (the **extraneous variable**) causes variation in the other two variables.

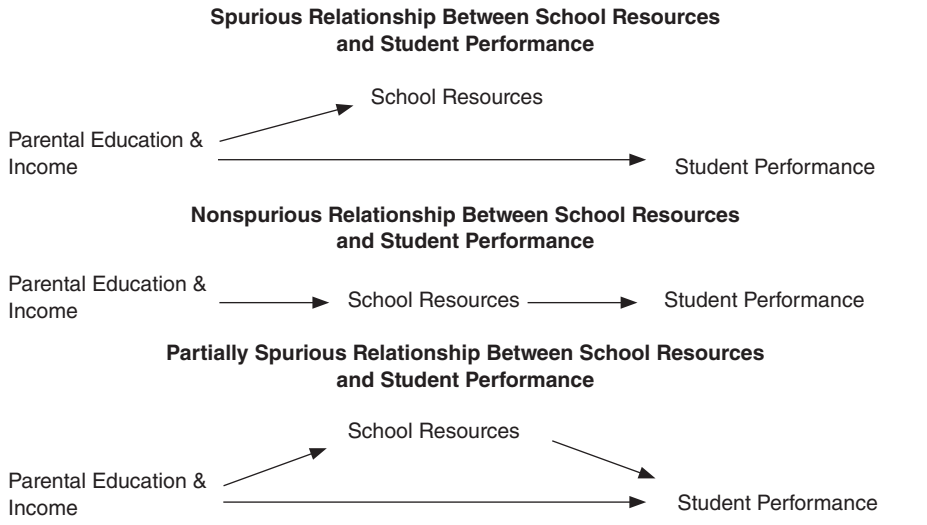
Always consider whether an association between two variables could have been caused by something other than the presumed independent variable—that is, by an extraneous variable. An extraneous variable may create a **spurious relationship** rather than a causal one. If we hypothesize that school resources (independent variable) influence student performance (dependent variable), for example, and measure them, we are likely to find a positive association between them. However, the association may result from the fact that parents with more education and income are more likely to send their children to schools with more resources but are also more likely to provide supports at home for their children that lead to higher levels of student performance. There could be no direct effect of school resources on student performance, even though there is an association between the level of resources at the school students attend and their academic performance (panel 1 in Figure 6.13). This would mean that the association of school resources with student performance is spurious, due to the extraneous variable (in terms of the original hypothesis) of parental education and income.

Alternatively, higher levels of parental education and income could increase the likelihood that students will attend schools with more resources, which in turn leads to higher levels of student performance (panel 2 in Figure 6.13). In this case, the association of parental education and income is not spurious. Instead, higher levels of school resources explain how parental education and income influence student performance.

Or perhaps higher levels of parental education and income result directly in higher levels of student performance as well as being associated with more school resources—and the resources in turn improve student performance (panel 3 in Figure 6.13). In this case, the association of school resources with student performance would be partially spurious because some of that association is because of the effect of parental education and income on student performance as well as on school resources.

We can't be sure if the association between school resources and student performance is spurious if our research design doesn't allow us to distinguish these different possibilities.

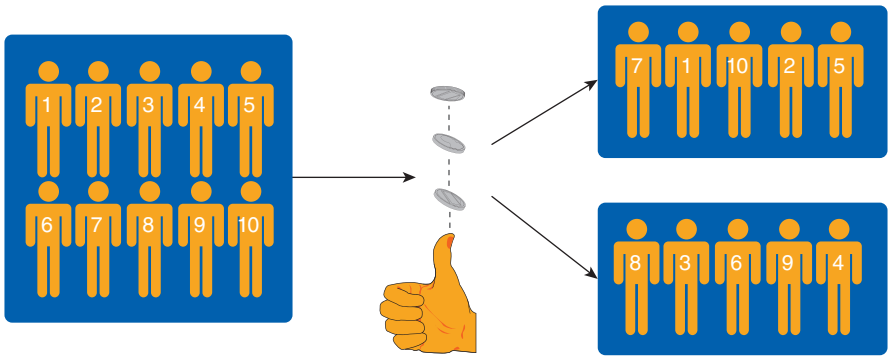
FIGURE 6.13 ■ Spurious, Nonspurious, and Partially Spurious Relationships


Spurious Relationship Between School Resources and Student Performance
School Resources
Parental Education & Income
Student Performance
Nonspurious Relationship Between School Resources and Student Performance
Parental Education & Income
School Resources
Student Performance
Partially Spurious Relationship Between School Resources and Student Performance
School Resources
Parental Education & Income
Student Performance

Randomization

Experimental design uses a technique called **randomization** to reduce the likelihood that an association between a hypothesized independent variable and a dependent variable is spurious. Lauren Donais and her coauthors' (2019) study of the impact of gender-based violence prevention workshops used randomization to reduce the risk of spuriousness. FEW classes were assigned randomly to either offer the VAWPP workshop after the students completed the questionnaire about their attitudes (the control condition) or before they completed the questionnaire (the treatment condition). That is, the classes students took were assigned randomly to one of the two categories of the independent variable (exposure to the VAWPP workshop). That would be like tossing a coin to decide which classes received the intervention (see Figure 6.14).

FIGURE 6.14 ■ Random Assignment to One of Two Groups



Random assignment ensures that neither the level of agreement with rape myths nor other student characteristics or attitudes could influence whether a class was in the treatment condition or the control condition. The greater the number of classes assigned randomly to the two conditions, the more likely the classes will be equivalent in all respects at the outset of the experiment. Whatever the preexisting sources of variation among the classes, these could not explain why the students in classes that had participated in the VAWPP workshop became less accepting of rape myths, whereas the others didn't. This is why tests of medical drugs and treatments (and the COVID-19 vaccines) use randomized controlled trials: when health and even lives hang in the balance, conclusions about causality should rest on the strongest possible research foundation.

Statistical Control

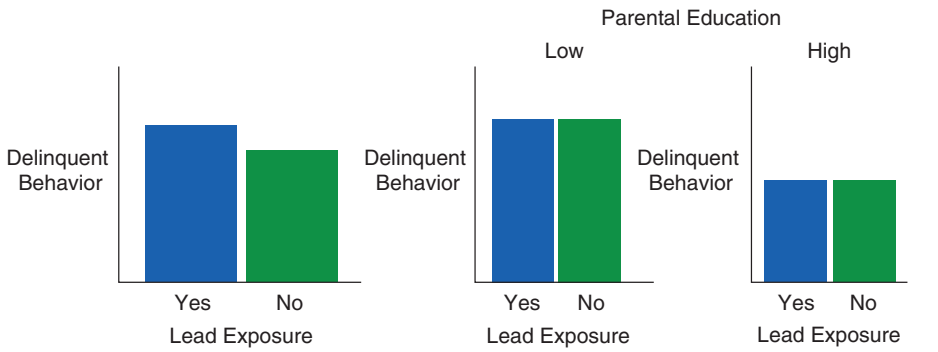
A nonexperimental study, such as Sampson and Winter's (2018), cannot assign cases randomly to comparison groups to minimize the risk of spurious effects. We can't randomly assign children to be exposed to lead or not, for ethical reasons—not to mention practical difficulties. Instead, nonexperimental researchers commonly use an alternative approach to try to achieve the criterion of nonspuriousness. The technique of **statistical control** allows researchers to determine whether the relationship between the independent and dependent variables still occurs while we hold constant the values of other variables. If it does, the relationship could not be caused by variation in these other variables.

Example of statistical control: In a different study, Sampson (1987) found a relationship between rates of family disruption and violent crime. He then classified cities by their level of joblessness (the control variable) and found the same relationship between the rates of family disruption and violent crime among cities with different levels of joblessness. Thus, the rate of joblessness could not have caused the association between family disruption and violent crime.

Sampson and Winter (2018) included in their study a test of whether the apparent effect of childhood lead exposure on delinquency was spurious because of the effect of background characteristics that might have been associated with both childhood lead exposure and adolescent delinquent behavior. These control variables were sex; race/ethnicity; immigrant generational status; and the primary caregiver's marital status, education level, and poverty. Sampson and Winter also controlled for age (the teenager's age at which their level of delinquency was measured) and neighborhood racial composition and neighborhood poverty. Even after controlling statistically for these other variables, childhood lead exposure still predicted adolescent delinquency. The hypothesized association was not spurious due to the influence of any of these control variables.

Figure 6.15 shows what would have happened if one of the control variables did create a spurious association between childhood lead exposure and adolescent delinquency. The first bar chart in Figure 6.15 shows the overall difference in adolescent delinquency for those who were exposed to lead as children and those who had not been exposed. The second bar chart shows that if we examine separately those with higher and lower primary caregiver's education, there is no difference in the level of delinquency between those who had and had not been exposed to lead as children. So when we statistically "control" for primary caregiver's education, we see that the amount of lead exposure makes no difference in relation to the level of delinquency. In this situation, controlling for the primary caregiver's educational level (bar charts on the right) "explains" the apparent effect of childhood lead exposure (bar chart on the left).

FIGURE 6.15 ■ The Use of Statistical Control to Reduce Spuriousness



This example illustrates the meaning of statistical control in order to test for spuriousness in an association between one variable we consider independent and another that it is hypothesized to influence. Most social scientists use more advanced statistics to control for multiple variables that might have caused a spurious association between two other variables, but the basic idea is the same. Using more advanced statistics allows researchers to control for multiple variables at one time, so in this way, they can account for many different potential sources of spuriousness.

What statistical control cannot do is control for variables that were not measured—including those the researcher didn't even think of including in their study. This is what makes experimental design a stronger method of controlling for possible sources of spuriousness: random assignment equates groups that differ in relation to the independent variable in terms of all other characteristics (within a margin of error).

Even though these two strategies for reducing the risk of spuriousness are very different, it's important to recognize that it can sometimes be useful to add the staistical control strategy to a study that uses a randomized experimental design. The Donais et al. (2019) study provides an example.

In their randomized experimental study of the effect of gender-based violence prevention workshops on student agreement with rape myths and other attitudes, Donais et al. (2019) also tested for possible spuriousness by controlling for several variables. Since randomization equates participants in the control and treatment conditions on all variables at the beginning of an experiment, controlling for initial characteristics in the analysis of outcomes is normally not necessary. However, Donais and her coauthors had randomized classes to the different conditions, not individual students. They realized that students might have enrolled in some systematic way in FYE classes that could have created some differences in the characteristics of students in different classes. They checked this out and found that the treated classes (those in which students had been exposed to a VAWPP workshop) tended to have somewhat fewer women, more athletes, and more Black and white students than those in the control classes.

To test whether these preexisting differences between students in the FYE classes with VAWPP and those without VAWPP could have influenced the difference in agreement with rape myths between these two conditions, Donais and her coauthors statistically controlled for these possible extraneous influences. They found that the differences in agreement with rape myths (and the other attitudes they measured) still occurred even after these statistical controls.

CAUTIONS FOR NOMOTHETIC CAUSAL EXPLANATIONS

We can strengthen our understanding of nomothetic causal connections and increase the likelihood of drawing causally valid conclusions by considering two cautions: the need to investigate causal mechanism and the need to consider the causal context. These two cautions are emphasized in the definition of idiographic causal explanation, with its attention to the sequence of events and the context in which they happen, but here I will limit my discussion to research oriented toward nomothetic causal explanations.

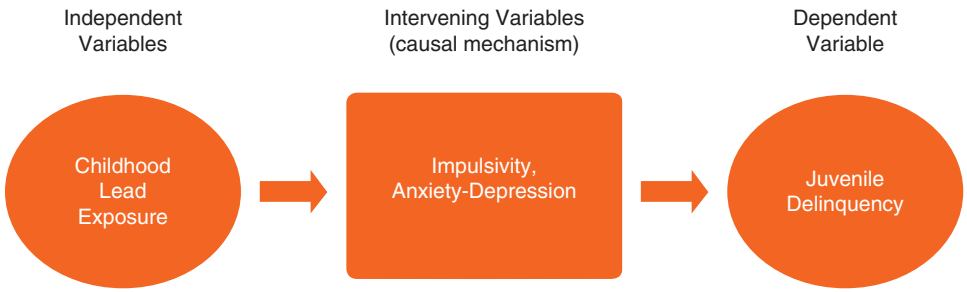
Mechanism

A causal **mechanism** is some process that creates the connection between variation in an independent variable and the variation in the dependent variable that the independent variable is hypothesized to cause (Cook and Campbell 1979:35; Marini and Singer 1988). Many social scientists (and scientists in other fields) argue that no nomothetic causal explanation is adequate until a causal mechanism is identified (Costner, 1989; Hedström and Swedberg 1998). In statistical analysis, variables that involve a mechanism are termed **mediators**.

Our confidence in causal conclusions based on nonexperimental research increases with identification of a causal mechanism (Shrout 2011:15–16). Such mechanisms help

us understand how variation in the independent variable results in variation in the dependent variable. For example, in their study of the effects of childhood lead exposure on juvenile delinquency, Sampson and Winter (2018) found that children exposed to lead became more impulsive, anxious, and depressed as adolescents, which may have explained why they also were more likely to become juvenile delinquents (see Figure 6.16). Impulsivity and anxiety-depression thus may have intervened in the relationship between childhood lead exposure and juvenile delinquency.

FIGURE 6.16 ■ Intervening Variables in Nonexperimental Research: Childhood Lead Exposure and Juvenile Delinquency



Source: Based on Sampson and Winter (2018).

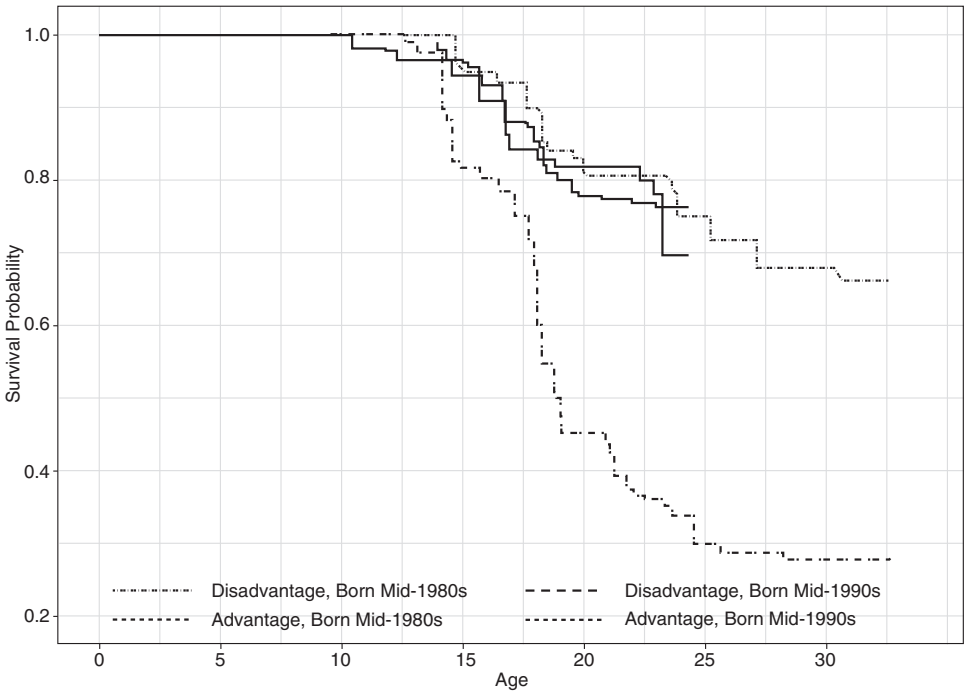
Context

Do the causal processes in which we are interested vary across neighborhoods? Between schools? Across regions? Over time? For different types of people? When relationships between variables differ across geographic units, such as counties or across other social settings, researchers say there is a **contextual effect**. Identification of the **context** in which a causal relationship occurs can help us understand that relationship. The changes in the crime rate with which we began this chapter differed for Black people and white people, for youth and adults, and in urban and rural areas (Ousey and Lee 2004:359–360). These contextual effects suggest that single-factor explanations about these changes are incorrect (Rosenfeld 2004:89). In statistical analysis, variables that identify contexts for the effects of other variables are termed **moderators**.

Neil and Sampson (2021) found support for contextual effects in their cohort study of predictors of likelihood of arrest among 1,057 individuals in three different birth cohorts. The likelihood of not having an arrest record plummeted after age 14 for disadvantaged youth born in the mid-1980s, but for those born in the mid-1990s and for those from advantaged backgrounds in both cohorts, the likelihood of not having an arrest record declined much more gradually in relation to age (see Figure 6.17). The results of this research suggest the importance of considering social context when examining criminal justice processes (see also Dannefer and Schutt 1982; Schutt and Dannefer, 1988).

Awareness of contextual differences can help make sense of discrepant findings from studies involving different locations, times, or types of people. Always remember the cause on which we focus in a particular research design may be only one among a set of interrelated factors required for the effect; when we account for context, we begin to specify these other factors (Hage and Meeker 1988; Papineau 1978).

FIGURE 6.17 ■ Contextual Effects of Disadvantage and Cohort on Association of Age and Delinquency



Source: Neil, Roland, and Sampson, Robert J. 2021. "The Birth Lottery of History: Arrest Over the Life Course of Multiple Cohorts Coming of Age, 1995-2018." *American Journal of Sociology,* 126(5):1156.

COMPARING RESEARCH DESIGNS

In summary, features of the different basic social research designs you will study in detail in the next chapters—experiments, surveys, and qualitative methods—provide distinct perspectives even when used to study the same social processes. Comparing subjects randomly assigned to a treatment and to a comparison group, asking standard questions of the members of a random sample, or observing while participating in a natural social setting involve markedly different decisions about measurement, causality, and generalizability. As you can see in Table 6.2, not one of these research designs can reasonably be graded as superior to the others with respect to all three dimensions of validity, so it is important to consider the tradeoffs when choosing one of these designs. Choosing among these three designs for a particular investigation also requires consideration of the research problem, opportunities and resources, prior research, philosophical commitments, and research goals.

Experimental designs are strongest for testing nomothetic causal hypotheses and are most appropriate for studies of treatment effects. Research questions that are believed to involve basic social psychological processes are most appealing for laboratory studies because the problem of generalizability is reduced. Random assignment reduces the possibility of preexisting differences between treatment and comparison groups to small, specifiable, chance levels; therefore, many of the variables that might create a spurious association are controlled. But despite this clear advantage, an experimental design requires a degree of control that cannot always be achieved outside the laboratory. It can be difficult to ensure in real-world settings that a treatment is

TABLE 6.2 ■ Comparison of Research Methods

Design	Measurement Validity	Generalizability	Type of Causal Assertions	Causal Validity
Experiments	+	-	Nomothetic	+
Surveys	+	+	Nomothetic	-/+[a]
Qualitative Methods	-/+[b]	-	Idiographic	-

a. Surveys are a weaker design for identifying causal effects compared to true experiments, although using statistical controls can strengthen causal arguments.

b. Reliability is low compared with surveys, and systematic evaluation of measurement validity is often not possible. However, direct observations or lengthy comments in qualitative interviews may lead to great confidence in the validity of measures.

delivered as intended and that other influences do not intrude. As a result, what appears to be a treatment effect or noneffect may be something else altogether. Field experiments thus require careful monitoring of the treatment process. Unfortunately, most field experiments also require more access arrangements and financial resources than can often be obtained.

Laboratory experiments permit much more control over conditions but at the cost of less generalizable findings. People must volunteer for most laboratory experiments, and so there is a good possibility that experimental subjects differ from those who do not volunteer. Ethical and practical constraints limit the types of treatments that can be studied experimentally (for example, you can't assign social class or race experimentally). The problem of generalizability in an experiment using volunteers lessens when the object of investigation is an orientation, behavior, or social process that is relatively invariant among people, but it is difficult to know which orientations, behaviors, or processes are so invariant. If a search of the research literature on the topic identifies many prior experimental studies, the results of these experiments will suggest the extent of variability in experimental effects and point to the unanswered questions about these effects.

Both surveys and experiments typically use standardized, quantitative measures of attitudes, behaviors, or social processes. Closed-ended questions are most common and are well suited for the reliable measurement of variables that have been studied in the past and whose meanings are well understood (see Chapter 4). Of course, surveys often include measures of many more variables than are included in an experiment (Chapter 8), but this feature is not inherent in either design. Phone surveys may be quite short, whereas some experiments can involve very lengthy sets of measures (see Chapter 7). The set of interview questions we used at baseline in the Boston housing study (Schutt 2011b) mentioned in Chapters 10 and 12, for example, required more than 10 hours to complete. The level of funding for a survey will often determine which type of survey is conducted and thus the length of the questionnaire.

Most social science surveys rely on random sampling for their selection of cases from some larger population, and this feature makes them preferable for descriptive research that seeks to develop generalizable findings (see Chapter 5). However, survey questionnaires can only measure what respondents are willing to report verbally; questionnaires may not be adequate for studying behaviors or attitudes that are regarded as socially unacceptable. Surveys are also often used to test hypothesized causal relationships. When variables that might create spurious relationships are included in the survey, they can be controlled statistically in the analysis and thus eliminated as rival causal influences.

CONCLUSIONS

Causation and the means for achieving causally valid conclusions in research is the last of the three legs on which the validity of research rests. In this chapter, you have learned about two alternative meanings of causation (nomothetic and idiographic). You have studied the three criteria and two cautions used to evaluate the extent to which particular research designs may achieve causally valid findings. You have learned how our ability to meet these criteria is shaped by research design features such as units of analysis, use of a cross-sectional or longitudinal design, and use of randomization or statistical control to deal with the problem of spuriousness. You have also seen why choosing an experimental or nonexperimental design has many consequences for how and how well we are able to meet nomothetic criteria for causation.

I should reemphasize that the results of any study are part of an always-changing body of empirical knowledge about social reality. Thus, our understandings of causal relationships are always partial. Researchers always wonder whether they have omitted some relevant variables from their controls, whether their experimental results would differ if the experiment were conducted in another setting, or whether they have overlooked a critical historical event. But by using consistent definitions of terms and maintaining clear standards for establishing the validity of research results—and by expecting the same of others who do research—social researchers can contribute to a growing body of knowledge that can reliably guide social policy and social understanding.

When you read the results of a social scientific study, you should now be able to evaluate critically the validity of the study's findings. If you plan to engage in social research, you should now be able to plan an approach that will lead to valid findings. And with a good understanding of the three dimensions of validity (measurement validity, generalizability, and causal validity) under your belt and with sensitivity also to the goal of *authenticity*, you are ready to focus on the major methods of data collection used by social scientists.

KEY TERMS

Association (p. 212)
Causal effect (idiographic perspective)
Causal effect (nomothetic perspective)
Ceteris paribus (p. 208)
Cohort (p. 206)
Context (p. 219)
Contextual effect (p. 219)
Counterfactual (p. 208)
Cross-sectional research design
Ecological fallacy (p. 198)
Emergence (p. 199)
Event-based design (cohort study)
Extraneous variable (p. 215)
Fixed-sample panel design (panel study)
Idiographic causal explanation (p. 209)
Longitudinal research design
Mechanism (p. 218)
Mediator
Moderator
Nomothetic causal explanation (p. 208)
Nonspuriousness (p. 214)
Random assignment (p. 216)
Randomization (p. 216)
Reductionist fallacy (reductionism)
Repeated cross-sectional design (trend study)
Spurious relationship (p. 215)
Statistical control (p. 216)
Subject fatigue (p. 206)
Time order (p. 200)
Units of analysis (p. 196)
Units of observation (p. 197)

HIGHLIGHTS

LO 6.1 Explain individual and group units of analysis and the role they play in the ecological and reductionist fallacies.

- In most sociological and psychological studies, the units of analysis are individuals. But in some studies, the units of analysis may instead be groups of some sort, such as families, schools, work organizations, towns, states, or countries.
- Invalid conclusions about causality may occur when relationships between variables measured at the group level are assumed to apply at the individual level (the ecological fallacy).
- Invalid conclusions about causality may occur when relationships between variables measured at the level of individuals are assumed to apply at the group level (the reductionist fallacy).
- Nonetheless, many research questions point to relationships at multiple levels and so may profitably be investigated at multiple units of analysis.

LO 6.2 Distinguish the three different longitudinal designs.

- Longitudinal designs vary in whether the same people are measured at different times, how the population of interest is defined, and how frequently follow-up measurements are taken.
- Repeated cross-sectional design (trend study): A type of longitudinal study in which data are collected at two or more points in time from different samples of the same population. This design is appropriate when the goal is to determine whether a population has changed over time.
- Fixed-sample panel design (panel study): A type of longitudinal study in which data are collected from the same individuals—the panel—at two or more points in time. This design is well suited for the testing of causal hypotheses, but it can be difficult to carry out successfully because of its expense as well as subject attrition and fatigue.
- Event-based design (cohort study): A type of longitudinal study in which the follow-up samples (at one or more times) are selected from the same cohort—people who all have experienced a similar event or a common starting point. An event-based design can be a type of repeated cross-sectional design or a type of panel design. Event-based research can improve identification of causal effects compared with cross-sectional designs.

LO 6.3 Illustrate an example of a nomothetic and of an idiographic causal explanation.

- A nomothetic causal explanation is one involving the belief that variation in an independent variable will be followed by variation in the dependent variable when all other things are equal (ceteris paribus). The value of cases on the dependent variable is measured after they have been exposed to variation in an independent variable. This measurement is compared with what the value of cases on the dependent variable would have been if they had not been exposed to the variation in the independent variable (the counterfactual). The validity of nomothetic causal conclusions rests on how closely the comparison group comes to the ideal counterfactual.
  - *Example of a nomothetic causal effect:* Individuals arrested for domestic assault tend to commit fewer subsequent assaults than do similar individuals who are accused in the same circumstances but not arrested.

- An idiographic causal explanation deals with the sequence of events that led to a particular outcome. In this type of causal explanation, individual events or the behaviors of individuals are explained with a series of related, prior events. It is more likely to be used for causal explanations in qualitative research.
  - *Example of an idiographic causal effect:* An individual is neglected by her parents but has a supportive grandparent. She comes to distrust others, has trouble in school, is unable to keep a job, and eventually becomes homeless. She subsequently develops a supportive relationship with a shelter case manager, who helps her find a job and regain her housing (based on Hirsch 1989).

LO 6.4 List the three criteria for establishing a nomothetic causal relationship.

- From a nomothetic perspective, three criteria are generally viewed as necessary for identifying a causal relationship: (1) association between the two variables, (2) proper time order, and (3) nonspuriousness of the association.
  - **Association:** Variation in one variable is related to variation in another variable
  - **Proper time order:** variation in the dependent variable occurred after the variation in the independent variable
  - **Nonspuriousness:** a relationship between two variables is not caused by variation in a third variable

LO 6.5 Recommend two cautions a researcher should keep in mind to improve understanding of a causal connection.

- We can strengthen our understanding of nomothetic causal connections and increase the likelihood of drawing causally valid conclusions by considering two cautions: the need to investigate causal mechanism and the need to consider the causal context.
  - **The need to investigate causal mechanism:** if we have not identified a causal mechanism, we do not understand fully *why* a causal connection exists. In statistical analysis, variables that involve a mechanism are termed mediators.
  - **The need to specify the context in which a causal effect occurs:** by understanding *when* or *under what conditions* the causal effect occurs, we will understand better what that effect *is*. In statistical analysis, variables that identify contexts for the effects of other variables are termed moderators.

LO 6.6 Appraise the strengths and weaknesses of employing an experimental design as compared to other methods.

- Strengths
  - Laboratory experiments permit much more control over conditions.
  - Experimental designs are strongest for testing nomothetic causal hypotheses and are most appropriate for studies of treatment effects.
  - Random assignment reduces the possibility of preexisting differences between treatment and comparison groups to small, specifiable, chance levels; therefore, many of the variables that might create a spurious association are controlled.
- Weaknesses:
  - An experimental design requires a degree of control that cannot always be achieved outside the laboratory (i.e., in field experiments).
  - Experimental designs produce less generalizable findings. People must volunteer for most laboratory experiments, and so there is a good possibility that experimental subjects differ from those who do not volunteer.

DISCUSSION QUESTIONS

1. There's a lot of "sound and fury" in the social science literature about units of analysis and levels of explanation. Some social researchers may call another a *reductionist* if the latter explains a problem such as substance abuse as caused by "lack of self-control." The idea is that the behavior requires consideration of social structure—a group level of analysis rather than an individual level of analysis. Another researcher may be said to commit an *ecological fallacy* if she assumes that group-level characteristics explain behavior at the individual level (such as saying that "immigrants are more likely to commit crime" because the *neighborhoods* with higher proportions of immigrants have higher crime rates). Do you favor causal explanations at the individual or the group (or social structural) level? If you were forced to mark on a scale from 0 to 100 the percentage of crime that results from problems with individuals rather than from problems with the settings in which they live, where would you make your mark? Explain your decision.

2. Researchers often try to figure out how people have changed over time by conducting a cross-sectional survey of people of different ages. The idea is that if people who are in their 60s tend to be happier than people who are in their 20s, it is because people tend to "become happier" as they age. But maybe people who are in their 60s now were just as happy when they were in their 20s, and people in their 20s now will be just as unhappy when they are in their 60s. (That's called a *cohort effect.*) We can't be sure unless we conduct a panel or cohort study (survey the same people at different ages). What, in your experience, are the major differences between the generations today in social attitudes and behaviors? Which would you attribute to changes as people age and which to differences between cohorts in what they have experienced (such as common orientations among millennials)? Explain your reasoning.

3. The chapter begins with some alternative explanations for recent changes in the crime rate. Which of the explanations make the most sense to you? Why? How could you learn more about the effect on crime of one of the "causes" you have identified in a laboratory experiment? What type of study could you conduct in the community to assess its causal impact?

4. This chapter discusses both experimental and nonexperimental approaches to identifying causes. What are the advantages and disadvantages of both approaches for addressing each of the three criteria and two cautions identified for causal explanations?

5. Construct an idiographic causal explanation for a recent historical or personal event. For example, what was the sequence of events that led to the outcome of the 2020 U.S. presidential election? What was the sequence of events that led to the shortage of toilet paper during the COVID-19 pandemic? (I know, you thought this would be easy.)

6. Randomization is a key feature of experimental designs that are often used to investigate the efficacy of new treatments for serious and often incurable, terminal diseases. What ethical issues do these techniques raise in studies of experimental treatments for incurable, terminal diseases? Would you make an ethical argument that there are situations when it is *more* ethical to use random assignment than usual procedures for deciding whether patients receive a new treatment?

SPSS EXERCISES

We can use the GSS2020 data to learn how causal hypotheses can be evaluated with nonexperimental data.

1. Specify four hypotheses in which cappun_2 is the dependent variable and the independent variable is also measured with a question in the 2020 GSS. The independent variables should have no more than 10 valid values (check the variable list).
   - **a.** Inspect the frequency distributions of each independent variable in your hypotheses. If it appears that one has little valid data or was coded with more than 10 categories, substitute another independent variable.
   - **b.** Generate cross-tabulations that show the association between cappun_2 and each of the independent variables. Make sure that cappun_2 is the row variable and that you select "Column Percents."
   - **c.** Does support for capital punishment vary across the categories of any of the independent variables? By how much? Would you conclude that there is an association, as hypothesized, for any pairs of variables?
   - **d.** Might one of the associations you have just identified be spurious because of the effect of a third variable? What might such an extraneous variable be? Look through the variable list and find a variable that might play this role. If you can't think of any possible extraneous variables or if you didn't find an association in support of any of your hypotheses, try this: Examine the association between cappun_2 and wrkstat_2. In the next step, control for sex (gender).

The idea is that there is an association between work status and support for capital punishment that might be spurious because of the effect of sex (gender). Proceed with the following steps:

- **i.** Select Analyze/Descriptive statistics/Crosstabs.
- **ii.** In the Crosstabs window, highlight cappun_2 and then click the right arrow to move it into Rows, Move wrkstat_2 into Columns and sex_2 into Layer 1 of 1.
- **iii.** Select Cells/Percentages Column/Continue/OK.

Is the association between employment status and support for capital punishment affected by gender? Do you conclude that the association between cappun_2 and wrkstat_2 seems to be spurious because of the effect of sex_2?

2. Does the association between support for capital punishment and any of your independent variables vary with social context? Test the association between attitude toward African Americans (helpblk_2) and cappun_2. Follow the procedures in SPSS Exercise 1d, but click helpblk_2 into columns and region_2 into Layer 1 of 1. (You must first return the variables used previously to the variables list.) Take a while to study this complex three variable table. Does the association between cappun_2 and helpblk_2 vary with region? How would you interpret this finding?

3. Now, how about the influence of an astrological sign on support for capital punishment? Create a cross-tabulation in which zodiac_2 is the independent (column) variable and cappun_2 is the dependent (row) variable (with column percents). What do you make of the results?

SECTION III

BASIC SOCIAL RESEARCH DESIGNS

Chapter 7	Experiments	228
Chapter 8	Survey Research	260
Chapter 9	Quantitative Data Analysis	301
Chapter 10	Qualitative Methods	349
Chapter 11	Qualitative Data Analysis	387

©iStockphoto.com/LaylaBird

7 EXPERIMENTS

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

7.1 Identify the essential components of a true experimental design

7.2 Assess the advantages and disadvantages of using a quasi-experimental design

7.3 State the features of research design that can reduce threats to internal validity

7.4 Explain why generalizability can be such a challenge in experimental research

7.5 Critique the most distinctive ethical challenges in experimental research

Whether or not you are already in the workforce, you'll probably be applying for a new position at some point after you earn your degree. Have you begun to think about who you'll ask to serve as a reference on your job applications? Are you thinking of applying for a job at an organization where you know someone you could ask to refer you? Doesn't that seem like a good idea? Will it matter if you're Black or white? If the employee who refers you is Black or white? How will you ever know?

These are the type of questions that sociologist Fabiana Silva (2018), at the University of Michigan's Gerald R. Ford School of Public Policy considered in her research, subsequently published in *Social Forces* as "The Strength of Whites' Ties: How Employers Reward the Referrals of Black and White Jobseekers." Her focus was not on what might happen to you or your friends, but on the extent to which people with hiring responsibilities in their workplace treat referrals for Black and white applicants the same. She knew she couldn't just ask job applicants about their experiences because this wouldn't reveal the basis for the employers' evaluations of the applicants. It wouldn't be a good idea to rely on employers' reports since she couldn't be sure if the employers would reveal how they treat Black and white applicants. Instead, Dr. Silva decided to conduct an experiment.

You learned in Chapter 6 that basic features of experimental designs help to identify causal effects. The study by Fabiana Silva and other experimental studies illustrate specific types of experimental designs that will multiply what you learned from the study summarized in Chapter 6 by Lauren Donais and her co-authors' (2019) about the effect of gender-based violence prevention workshops. This chapter distinguishes different types of experimental design (both true experiments and quasi-experiments). Next, it reviews how well particular designs establish causally valid conclusions and achieve generalizable results. Finally, the chapter considers ethical issues that should be given particular attention when considering experimental research.

Experimental research by sociologists, social psychologists, and psychologists as well as other social scientists has made important contributions to understanding the social world since the middle years of the 20th century. Laboratory experiments by Harvard's Solomon Asch (1958) showed that individuals would yield to social pressure in their own judgments about the length of lines—even when they knew their perceptions were correct. Experiments by Berger et al. (1972) made it clear that students quickly deferred in conversation to other students of higher status. Bargh et al. (1996) found that having

students read words that included some polite, neutral, or rude terms had a "priming effect" that led them immediately after to behave in a corresponding manner. Responding to concerns about the effect of economic hardship on crime, Berk, Lenihan, and Rossi (1980) conducted an experimental test "in the field" of the impact of providing financial subsidies to incarcerated persons who were released for one to two years. (It had no impact on their likelihood of reoffending.)

By the time you finish the chapter, you should understand why experimentation is often the method of choice when the research question is about the effect of a treatment or some other variable whose values can be manipulated by the researcher—whether the treatments are vaccines for SARS-CoV-2, programs to reduce gun violence, or something else. You will also appreciate why experimental methods cannot be used to study many of the research questions social scientists investigate.

TRUE EXPERIMENTS

The value of experimental methods is clearest when we focus on what are termed *true experiments*. **True experiments** must have at least three features:

1. Two groups (in the simplest case, an experimental and a control group)
2. Variation in the independent variable before assessment of change in the dependent variable
3. Random assignment to the two (or more) comparison groups

The combination of these features permits us to have much greater confidence in the validity of causal conclusions than is possible in other research designs. As you learned in Chapter 6, two more features further enhance our confidence in the validity of an experiment's findings:

1. Identification of the causal mechanism
2. Control over the context of an experiment

You will learn more about each of these key features of experimental design as you review experimental studies about social processes. I use simple diagrams to help describe and compare the experiments' designs. These diagrams also show how well suited any experiment is to identifying causal relationships, by indicating whether the experiment has a comparison group, a pretest and a posttest, and randomization.

Experimental and Comparison Groups

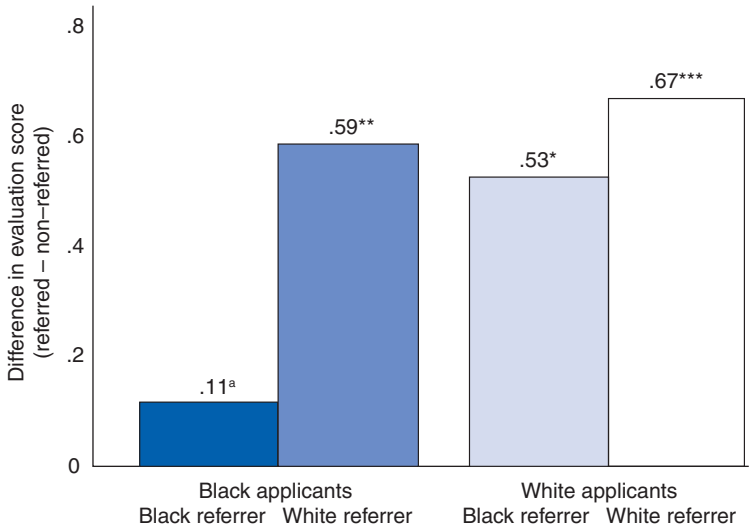
True experiments must have at least one **experimental group** (subjects who receive some treatment) and at least one **comparison group** (subjects with whom the experimental group can be compared). The comparison group differs from the experimental group in one or more independent variables whose effects are being tested. In other words, variation in the independent variable determines the difference between the experimental and comparison groups.

In many experiments, the independent variable indicates the presence or absence of something, such as receiving a treatment program or not receiving it. In these experiments, the comparison group, consisting of the subjects who do not receive the treatment, is termed a **control group**. You learned in Chapter 6 that an experiment can have more than two groups. There can be several treatment groups, corresponding to different values of the independent variable, and several comparison groups—including a control group that receives no treatment.

Silva (2018) used four comparison groups in her experiment about how the race of employee referrals affects decisions by white individuals with hiring responsibilities about Black and white job applicants. She recruited participants for the experiment through Amazon's Mechanical Turk (MTurk) service, an online platform in which people agree to perform tasks for a small per-task fee. The result was a sample of 226 white, non-Hispanic workers who had indicated they had hiring responsibilities in their own workplaces and agreed to provide their opinion about applicants for a position at a retail company. The participants read a brief job description, an [fictitious] applicant's resume, a [fictitious] referral form submitted for that applicant, and then a resume and referral form for a second [fictitious] applicant.

The race of job applicants and those who submitted a referral form was not stated explicitly but was indicated through racially distinct names, such as Charlie to suggest a white applicant and Jermaine to suggest a Black applicant. (A check after the experiment indicated respondents matched the names with the expected race.) The study results indicated that Black applicants received almost no benefit in a hiring agents' evaluations when they had been referred by a Black employee, but both Black and white applicants received a more positive evaluation if they were referred by a white employee (see Figure 7.1). Ironically, white applicants also benefitted from referrals by Black employees, even though Black applicants did not receive this benefit.

FIGURE 7.1 ■ Effect of Referral Status on Evaluation Score, by Experimental Condition



Note: Statistical tests based on linear regressions of referral status on the evaluation score, with standard errors clustered at the respondent level. The evaluation score is standardized: regression coefficient (two-tailed tests); *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ (two-tailed tests). See complete article for details.

Source: Silva, Fabiana. 2018. "The Strength of Whites' Ties: How Employers Reward the Referrals of Black and White Jobseekers." *Social Forces* 97(2):741–768.

Pretest and Posttest Measures

All true experiments have a **posttest** —that is, the measurement of the outcome in both groups after the experimental group has received the treatment. Many true experiments also have a **pretest** that measures the dependent variable before the experimental intervention. A pretest is the same as a posttest, except that it is administered at a different time. Pretest scores permit a direct measure of how much the experimental and comparison groups changed during the experiment, from before the experimental treatment occurred until after it was completed.

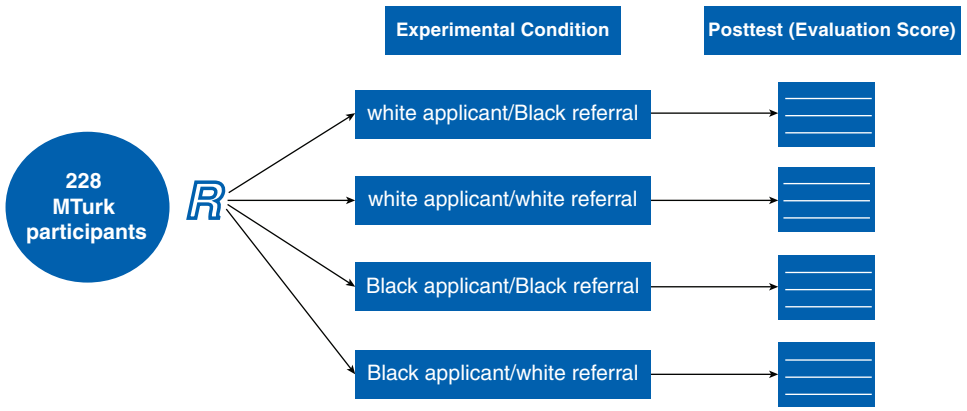
When a pretest is used, it also allows the researcher to verify that randomization was successful (chance factors did not lead to an initial difference between the groups). In addition, by identifying subjects' initial scores on the dependent variable, a pretest provides a more complete picture of the conditions in which the intervention had (or didn't have) an effect (Mohr 1992:46–48). A randomized experimental design with a pretest and posttest is termed a **randomized comparative change design** or a *pretest–posttest control group design.*

An experiment may have multiple posttests and perhaps even multiple pretests. Multiple posttests can identify just when the treatment has its effect and for how long. This is particularly important when treatments are delivered over time (Rossi and Freeman, 1989:289–290).

Strictly speaking, however, a true experiment does not require a pretest. When researchers use random assignment to the experimental and comparison groups, the groups' initial scores on the dependent variable and on all other variables are likely to be similar. Any difference in outcome between the experimental and comparison groups is therefore likely to result from the intervention (or from other processes occurring during the experiment), and the likelihood of a difference occurring just based on chance can be calculated. This is fortunate because the dependent variable in some experiments cannot be measured in a pretest. This was the case in Silva's experiment (2018) since her dependent variable—evaluation scores for the job applicants—could not be measured before the decision makers had learned about the applicants by reading their referral forms and other evaluation materials.

Figure 7.2 diagrams the Silva study. The labels indicate the study used a *posttest-only control group design,* also called the **randomized comparative posttest design**. You'll also learn later in this chapter that there can be a disadvantage to having a pretest, even when it is possible to do so: The act of taking the pretest can itself cause participants to change.

FIGURE 7.2 ■ Diagram of Silva Experiment on the Effect of Race of Job Applicants and Referral Sources



Source: Based on Silva (2018:748–751).

Randomization

Randomization, or random assignment, is what makes the comparison group in a true experiment like Donais, Simonsen, and Simonsen (2019) and Silva's (2018) such a powerful tool for identifying the effects of the treatment. A randomized comparison group can provide a good estimate of the counterfactual—the outcome that would have occurred if the subjects who were exposed to the treatment had not been exposed but otherwise had had the same experiences (Mohr 1992; Rossi and Freeman 1989:229). A researcher cannot determine for sure what the unique effects of a treatment are if the comparison group differs from the experimental group in any way other than not receiving the treatment.

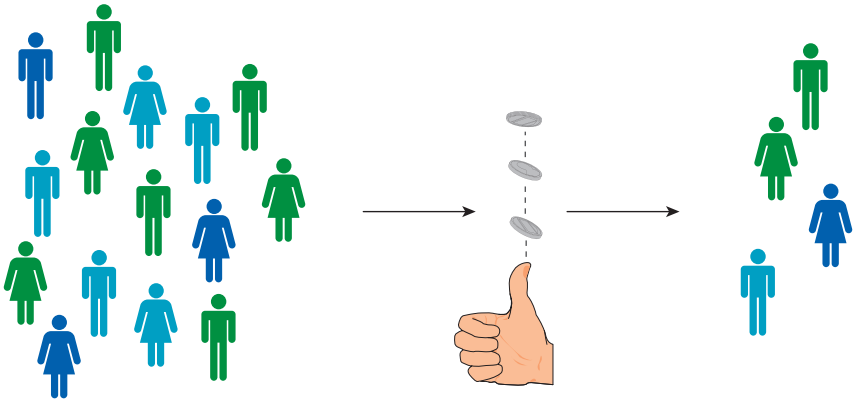
Assigning subjects randomly to the experimental and comparison groups ensures that systematic bias does not affect the assignment of subjects to groups. Of course, random assignment cannot guarantee the groups are perfectly identical at the start of the experiment. Randomization removes bias from the assignment process but only by relying on chance, which itself can result in some intergroup differences (Bloom 2008:116). Fortunately, researchers can use statistical methods to determine the odds of ending up with groups that differ very much because of chance, and these odds are low even for groups of moderate size. The larger the group, the less likely it is that even modest differences will occur based on chance and the more possible it becomes to draw conclusions about causal effects from relatively small differences in the outcome.

Note that the random assignment of subjects to experimental and comparison groups is not the same as a random sampling of individuals from some larger population (see Figure 7.3). In fact, random assignment (randomization) does not help at all to ensure the research subjects are representative of some larger population; instead, representativeness is the goal of random sampling. What random assignment does—create two (or more) equivalent groups—is useful for maximizing the likelihood of internal validity, not generalizability (Bloom 2008:116).

Matching is another procedure used to equate experimental and comparison groups, but by itself, it is a poor substitute for randomization. Matching of individuals in a treatment group with those in a comparison group might involve pairing persons based on similarity of gender,

FIGURE 7.3 ■ Random Sampling Versus Random Assignment

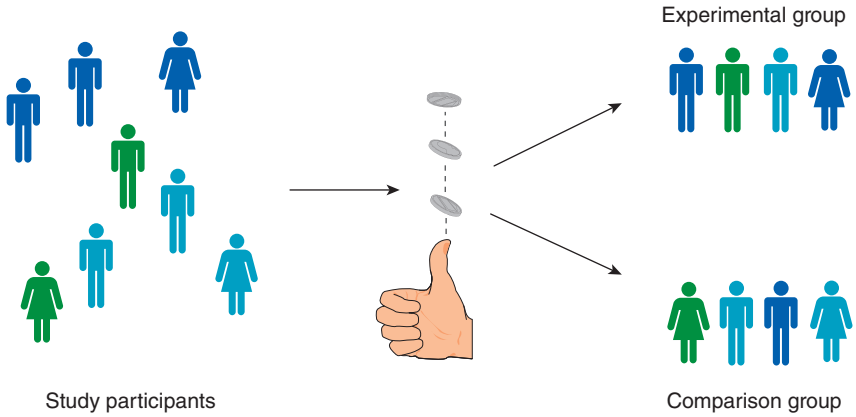
Random sampling (a tool for maximizing generalizability):
Individuals are randomly selected from a population to participate in a study.



age, year in school, or some other characteristic. The basic problem is that, as a practical matter, individuals can be matched on only a few characteristics; as a result, unmatched differences between the experimental and comparison groups may still influence outcomes. When matching is used as a substitute for random assignment, the research becomes *quasi*-experimental instead of being a true experiment. However, matching combined with randomization, also called *blocking*, can help to reduce the possibility of differences resulting from chance (Bloom 2008:124). For example, if individuals are matched in gender and age and then the members of each matched pair are assigned randomly to the experimental and comparison groups, the possibility of outcome differences because of chance differences in the gender and age composition of the groups is eliminated (see Figure 7.4).

FIGURE 7.4 ■ Experimental Design Combining Matching and Random Assignment

Random assignment, or randomization (a tool for maximizing internal validity): Individuals who are to participate in a study are randomly divided into an experimental group and a comparison group.



The **Solomon four-group design** is a true experimental design that combines a randomized comparative change design (the pretest–posttest control group design) with the randomized comparative posttest design (posttest-only control group design). This design allows comparison of the effect of the independent variable on groups that had a pretest with its effect on groups that have not had a pretest (see Table 7.2 , later in this chapter). Whenever there is reason to think that taking the pretest may itself influence how participants react to the treatment, the Solomon four-group design should be considered. If the pretest had such an effect, the difference in outcome scores between the experimental and comparison groups will be different for subjects who took the pretest (the pretest–posttest design) compared with those who did not (the posttest-only design).

Limitations of True Experimental Designs

The distinguishing features of true experiments—experimental and comparison groups, pretests (which are not always used) and posttests, and randomization—can be implemented most easily in laboratory settings with participants such as students who are available for such experiments. For this reason, true experimental designs are used most often in social psychological

experiments that focus on research questions about reactions to conditions that can easily be created in laboratories on college campuses. However, this focus on college students in laboratory settings raises the question of whether findings can be generalized to other populations and settings. This problem of generalizability is the biggest limitation of true experimental designs, and so I will return to it later in the chapter. You will read examples of experiments conducted outside of the laboratory in this chapter's section on factorial surveys and in Chapter 13, on evaluation and policy research.

The potential for the pretest to influence participants' reactions to the experimental treatment is also a limitation of experimental designs, but it is one that can be solved by researchers who have sufficient funds and time to double the number of participants in the experiment and use a Solomon four-group design.

The criteria for true experimental designs also do not help researchers identify the mechanisms by which treatments have their effects. In fact, this question of causal mechanisms often is not addressed in experimental research. The hypothesis test itself does not require any analysis of mechanism, and if the experiment was conducted under carefully controlled conditions during a limited time span, the causal effect (if any) may seem to be quite direct. But attention to causal mechanisms can augment experimental findings. For example, evaluation researchers often focus attention on the mechanisms by which a social program has its effect (Mohr 1992:25–27; Scriven 1972a). The goal is to measure the intermediate steps leading to the change that is the program's primary focus.

True experimental designs also do not guarantee the researcher has been able to maintain control over the conditions to which subjects are exposed after they are assigned to the experimental and comparison groups. If these conditions begin to differ, the variation between the experimental and comparison groups will not be what was intended. Such unintended variation is often not much of a problem in laboratory experiments, where the researcher has almost complete control over the conditions (and can ensure these conditions are nearly identical for both groups). But control over conditions can become a big concern for **field experiments**, experimental studies that are conducted in the field, in real-world settings. Silva's (2018) experimental study of responses to the race of job applicants and those who referred them was not conducted in the laboratory, but it also wasn't conducted in a real-world setting: the "Turkers" who participated expected to carry out tasks for a fee and are unlikely to have been affected by some general conditions around them. By contrast, in Sherman and Berk's (1984) field experiment about the police response to domestic violence (see Chapter 2), police officers did not adhere consistently to the requirement that they arrest suspects based on the random assignment protocol. As a result of this concern, the subsequent replications of the experiment in other cities reduced the discretion of individual police officers and used a more centralized procedure for assigning cases to the treatment and comparison conditions.

Summary: Causality in True Experiments

The study by Silva (2018) was a true experiment because it had at least one experimental and one comparison group to which subjects were randomly assigned. The researchers also compared variation in the dependent variables after variation in the independent variables (race of applicant and source of referral). Although Silva did not have a pretest score for the dependent variable she measured in the posttest, you have learned that a pretest is not required in a true experiment. Donais and her co-authors (2019) did not have a pretest in their experimental study of violence prevention workshops (Chapter 6), nor did Sherman and Berk's (1984) experimental study of the police response to domestic violence (Chapter 2).

Let's examine how well true experiments meet the criteria for identifying a nomothetic cause (introduced in Chapter 6).

Association Between the Hypothesized Independent and Dependent Variables

As you have seen, experiments can provide unambiguous evidence of association by comparing the distribution of the dependent variable (or its average value) between the experimental and comparison groups.

Time Order of Effects of One Variable on the Others

Unquestionably, the application materials describing applicants and sources of referrals were received before the MTurk "deciders" in the Silva study made their evaluations, the use of body-worn cameras preceded the measure of citizen complaints in the Braga et al. (2018) study (described in Chapter 2), and the violence prevention workshop in the Donais et al. (2019) study preceded the differential changes in rejection of rape myths between the experimental and comparison groups. In true experiments, randomization to the experimental and comparison groups equates the groups at the start of the experiment, so time order can be established by comparing posttest scores between the groups. However, experimental researchers include a pretest when possible so that equivalence of the groups at baseline can be confirmed and the amount of change can be compared between the experimental and comparison groups.

Nonspurious Relationships Between Variables

Nonspuriousness is difficult—some would say impossible—to establish in nonexperimental designs. The random assignment of subjects to experimental and comparison groups is what makes true experiments such powerful designs for testing causal hypotheses. Randomization controls for the host of possible extraneous influences that can create misleading, spurious relationships in both experimental and nonexperimental data. If we determine that a design has used randomization successfully, we can be much more confident in the resulting causal conclusions.

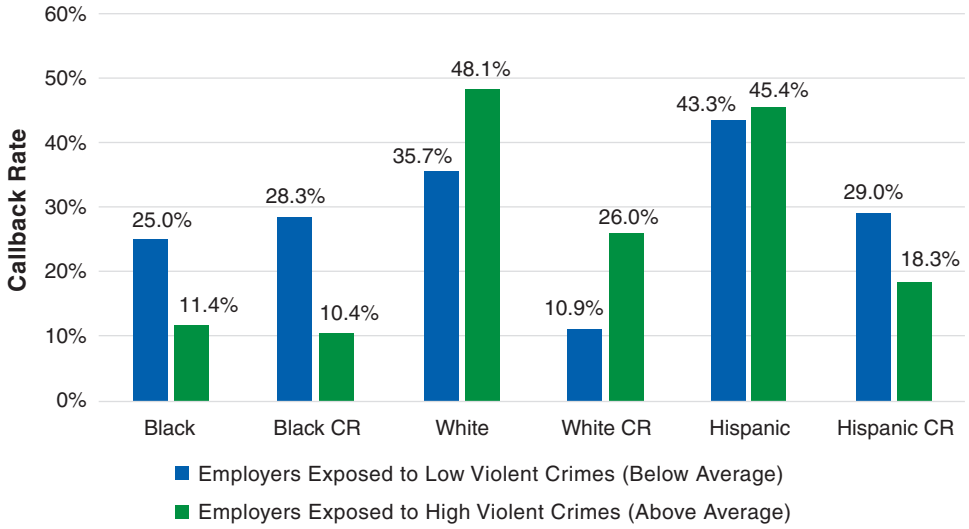
Mechanism That Creates the Causal Effect

The core features of a true experiment do not in themselves allow identification of causal mechanisms; as a result, there can be some ambiguity about how the independent variable influenced the dependent variable and the resulting causal conclusions (Bloom 2008:128). However, experimental researchers may add features to their experiments to focus on the causal mechanism that helps explain the effect of the independent variable. For example, Silva (2018)

Context in Which Change Occurs

Control over conditions is more feasible in many experimental designs than it is in nonexperimental designs. When participants in an experiment are exposed to the treatment in a laboratory, the researcher can ensure that conditions are the same for each participant and that they remain the same each time the experiment is repeated. However, when an experiment is conducted "in the field," this may not be possible. In a study of the treatment of Black job applicants with and without a criminal record for a position at one of 184 employers in Oakland, California, Mobasseri (2019) measured recent exposure to violent crime in the vicinity of the employer. Black job applicants were less likely to be called back for an interview if there had been violent crimes near the employer, whether the Black applicant had a criminal record or not. This effect of context did not occur with white or Hispanic job applicants (see Figure 7.5).

FIGURE 7.5 ■ Callback Rates, by Race and Exposure to Violent Crime Events



Source: Mobasseri, Sanaz. 2019. "Race, Place, and Crime: How Violent Crime Events Affect Employment Discrimination." *American Journal of Sociology* 125(1):91.

RESEARCH IN THE NEWS

Study Shows Which Firms Discriminate

In 2019, economists at the University of California, Berkeley, and the University of Chicago began a large discrimination audit by mailing 83,000 fake job applications for entry-level positions to 108 large U.S. companies. Applicant characteristics like age, sexual orientation, and work and school experience were varied at random, but similar applications were sent in pairs that differed in the applicants' names: One had a distinctive white name, such as Jake or Molly, and the other had a distinctive Black name, such as DeShawn or Imani.

On average, applications from candidates with a "Black name" got fewer callbacks than those from candidates with a "white name." However, the apparent discrimination was concentrated in particular firms. The most discriminatory firms tended to be less profitable—as predicted by economic theory that it is costly for firms to discriminate against productive workers by taking into account extraneous considerations like race—and the companies with more centralized personnel operations tended to discriminate less. There were no regional differences in the evidence of hiring discrimination.

For Further Thought?

1. What features make this study experimental?
2. What causal mechanisms might explain the causal effect of race in hiring decisions? How would you design a study to investigate these mechanisms?

News source: Porter, Eduardo. 2021. "Study Shows Which Firms Discriminate." *The New York Times*, July 30, B1, B6.

QUASI-EXPERIMENTS

Often, testing a hypothesis with a true experimental design is not feasible with the desired participants and in the desired setting. Such a test may be too costly or take too long to carry out, it may not be ethical to randomly assign subjects to the different conditions, or it may be too late to do so. In these situations, researchers may instead use *quasi-experimental* designs that retain several components of experimental design but do not randomly assign participants to different conditions.

A **quasi-experimental design** is one in which the comparison group is predetermined to be comparable with the treatment group in critical ways, such as being eligible for the same services or being in the same school cohort (Rossi and Freeman 1989:313). These research designs are *quasi*-experimental because subjects are not randomly assigned to the comparison and experimental groups. As a result, we cannot be as confident in the comparability of the groups as in true experimental designs. Nonetheless, to term a research design quasi-experimental, we must be sure the comparison groups meet specific criteria that help lessen the possibility of preexisting differences between groups.

I discuss here the two major types of quasi-experimental designs—nonequivalent control group designs and before-and-after designs—as well as a nonexperimental design that can be very similar to nonequivalent control group designs (other types can be found in Cook and Campbell 1979; Mohr 1992):

- **Nonequivalent control group designs** have experimental and comparison groups that are designated before the treatment occurs and are not created by random assignment.
- **Before-and-after designs** have a pretest and posttest but no comparison group. In other words, the subjects exposed to the treatment serve, at an earlier time, as their own controls.

Nonequivalent Control Group Designs

The nonequivalent control group design is the most common type of quasi-experimental design, also called a *differences-in-differences* design. In it, a comparison group is selected to be as comparable as possible to the treatment group. Two selection methods can be used: aggregate matching and individual matching.

Aggregate Matching

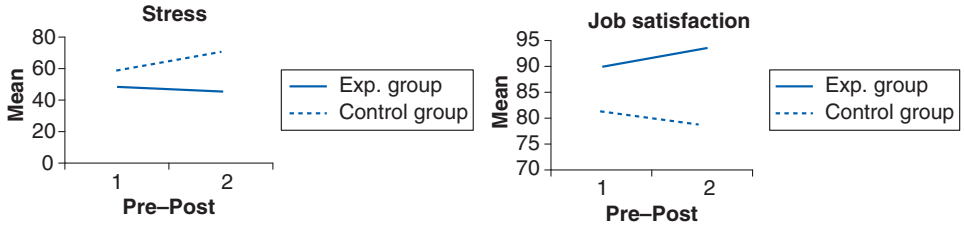
Once research moves outside of laboratories on college campuses and samples of available students, it becomes much harder to control what people do and what type of experiences they will have. When random assignment is not possible, an alternative approach to testing the effect of some treatment or other experience can be to find a comparison group that matches the treatment group in many ways but differs in exposure to the treatment. For example, a sociologist who hypothesizes that the experience of a disaster increases social solidarity might identify two towns that are similar in population characteristics but where one of the two towns experienced an unexpected disaster. If the populations in the two towns have similar distributions on key variables such as age, gender, income, and so on, but they differ after the disaster in their feelings of social solidarity, the sociologist might conclude a higher level of solidarity in the affected town resulted from the disaster. However, it is important in a nonequivalent control group design that

individuals have not been able to choose whether to join the group that had one experience or the other. If many people moved from the stricken town to the unaffected town right after the disaster, higher levels of post-disaster solidarity among residents in the stricken town could reflect the departure of people with feelings of low solidarity from that town rather than the effect of the disaster on increasing feelings of solidarity.

Innstrand, Geir, and Mykletun (2004) used a nonequivalent control group design with aggregate matching to study the effect of a new program in Norway to reduce stress among staff working with people with intellectual disabilities. The researchers chose two Norwegian municipalities that offered the same type of community residential care and used similar staff. The two municipalities were in different locations and had no formal communication with each other.

When their research began, Innstrand and colleagues asked staff in both municipalities to complete an anonymous survey that measured their stress, job satisfaction, and other outcomes. This was their pretest. The researchers then implemented the stress reduction program in one municipality. This was their experimental treatment. They distributed their survey again 10 months later. This was their posttest. Two of their primary findings appear in Figure 7.6. Levels of stress declined in the experimental group and increased in the control group, and job satisfaction increased in the experimental group and declined in the control group. It seemed the program had at least some of the effects that Innstrand and colleagues predicted.

FIGURE 7.6 ■ Findings of the Innstrand Quasi-Experimental Study



Source: Innstrand, Siw Tone, Geir Arild Espries, and Reidar Mykletun. 2004. "Job Stress, Burnout and Job Satisfaction: An Intervention Study of Staff Working With People With Intellectual Disabilities." *Journal of Applied Research in Intellectual Disabilities* 17:119–126. Copyright © 2004, John Wiley & Sons. Reprinted with permission.

Nonequivalent control group designs based on aggregate matching should always be considered when the goal is to compare effects of treatments or other experiences and a true experiment is not feasible. However, simply comparing outcome measures in two groups that offer different programs is not in itself a quasi-experiment. If individuals can choose which group to join, partly based on the program offered, then the groups will differ in preference for the treatment as well as in having had the treatment. When such selection bias is possible, the design is nonexperimental rather than quasi-experimental. More generally, the validity of this design depends on the adequacy of matching the comparison group with the treatment group (Cook and Wong 2008:151).

Individual Matching

In individual matching, individual cases in the treatment group are matched with similar individuals in the comparison group. In some situations, this can create a comparison group that is very similar to the experimental group, as when children in Head Start were matched with their siblings to estimate the effect of participation in Head Start (Currie and Thomas 1995:341).

However, in many studies, it is not possible to match in this way on the most important variables. Simply matching based on such readily available characteristics as gender or age will be of little value if these sociodemographic characteristics are not related to propensity to participate in the treatment group or extent of response to treatment (Cook and Wong 2008:153).

Variables chosen for matching affect the quality of a nonequivalent comparison group design (Cook and Wong 2008:154–155). Matching with comparison groups comprising twins, siblings, members of the same organization, or others who are very similar will often be the best choice.

The quality of a nonequivalent comparison group design can also be improved by the inclusion of several other features (Cook and Wong 2008:154–155):

- A pretest in both groups, so the starting point for the groups can be compared
- Identical procedures to measure outcomes in the two groups (same measure, same time of assessment)
- Several outcome measures reflecting the same overall causal hypothesis
- Investigating the process of selection into the two groups and controlling for the elements of this process that can be measured

Quasi-experimental designs with these features can result in estimates of effects that are very similar to those obtained with a randomized design (Cook and Wong 2008:159).

Before-and-After Designs

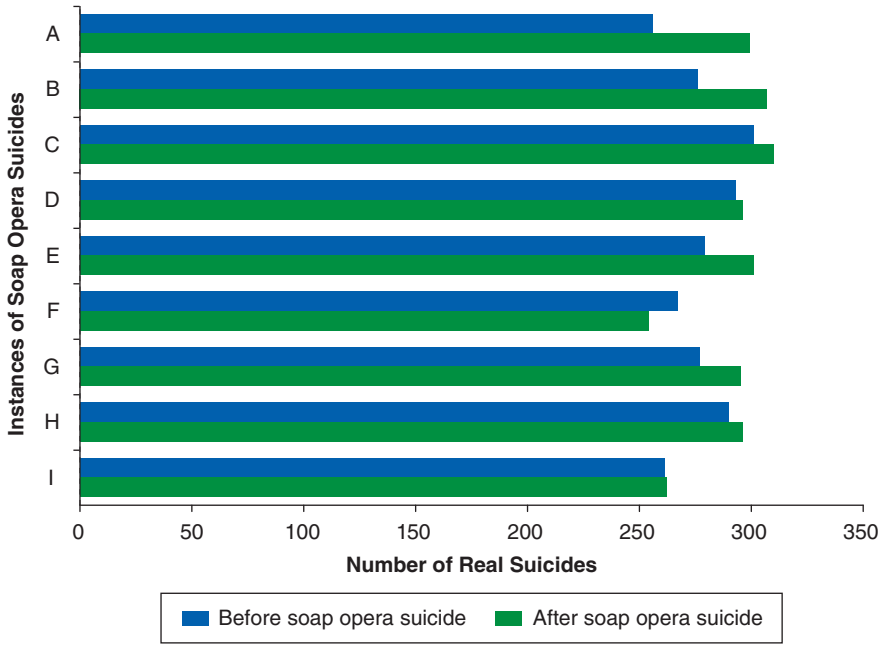
The common feature of before-and-after designs is the absence of a comparison group. Because all cases are exposed to the experimental treatment, the basis for comparison is provided by comparing the pretreatment with the posttreatment measures. These designs are thus useful for studies of interventions that are experienced by virtually every case in some population, including total coverage programs, such as Social Security or studies of the effect of a new management strategy in a single organization. The simplest type of before-and-after design is the fixed-sample panel design, with one pretest and one posttest (see Chapter 6).

David Phillips's (1982) study of the effect of TV soap opera suicides on the number of actual suicides in the United States illustrates a more powerful **multiple group before-and-after design**. In this design, several before-and-after comparisons are made involving the same variables but with different groups. Phillips identified 13 soap opera suicides in 1977 and then recorded the U.S. suicide rate in the weeks before and following each TV story. Because several suicides occurred in adjacent weeks, the analysis proceeded as if there had been nine soap opera suicides. In effect, the researcher had nine different before-and-after studies, one for each suicide story occurring in a unique week. In eight of these nine comparisons, deaths from suicide increased from the week before each soap opera suicide to the week after (see Figure 7.7).

Another type of before-and-after design involves multiple pretest and posttest observations of the same group. These may be **repeated measures panel designs**, which include several pretest and posttest observations, and **time series designs**, which include many (preferably 30 or more) such observations in both pretest and posttest periods. Repeated measures panel designs are stronger than simple before-and-after panel designs because they allow the researcher to study the process by which an intervention or treatment has an impact over time.

In a time series design, the trend in the dependent variable until the date of the intervention or event whose effect is being studied is compared with the trend in the dependent variable after the intervention. A substantial disparity between the preintervention trend and the

FIGURE 7.7 ■ Real Suicides and Soap Opera Suicides



Source: Adapted from Phillips (1982):1347. Reprinted with permission from the University of Chicago Press.

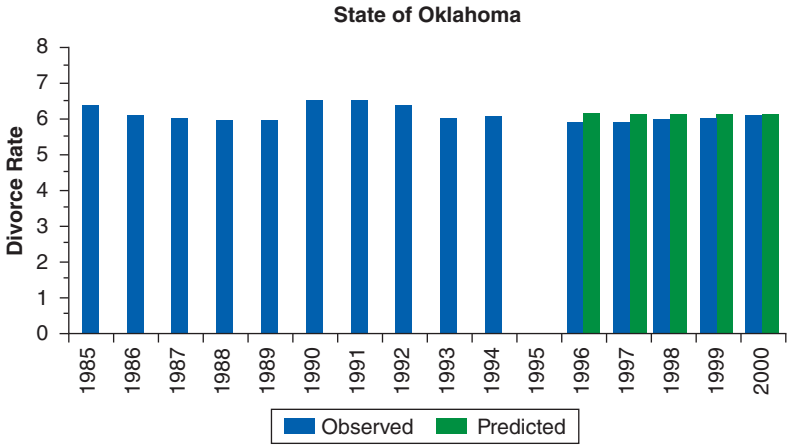
postintervention trend is evidence that the intervention or event had an impact (Rossi and Freeman 1989:260–261, 358–363).

Time series designs are particularly useful for studies of the impact of new laws or social programs that affect everyone and that are readily assessed by some ongoing measurement. For example, Nakonezny, Reddick, and Rodgers (2004) used a time series design to identify the impact of the Oklahoma City terrorist bombing in April 1995 on the divorce rate in Oklahoma. They hypothesized that people would be more likely to feel a need for support in the aftermath of such a terrifying event and thus be less likely to divorce. Nakonezny et al. first calculated the average rate of change in divorce rates in Oklahoma's 77 counties in the 10 years before the bombing and then projected these rates forward to the 5 years after the bombing. As they hypothesized, they found the actual divorce rate in the first years after the bombing was lower than the prebombing trend would have predicted, but this effect diminished to nothing by the year 2000 (see Figure 7.8).

The most powerful type of quasi-experimental design that can be considered a before-and-after design is the **regression–discontinuity design**. This type of design can be used if participants are assigned to treatment solely based on a cutoff score on some **assignment variable**. Students may be admitted to a special intensive course based on a test score, or persons may become eligible for a housing voucher based on their income. In these two situations, test score and personal income are assignment variables.

Researchers using a regression–discontinuity design plot the relationship (the regression line) between scores on the assignment variable and the outcome of interest for those who did not enter the treatment as well as for those who did enter the treatment. If there is a jump in the regression line at the cutoff score, it indicates an effect of the treatment. For example, Jalbert et al. (2010) studied the effect of reduced caseload size and intensive supervision on probation

FIGURE 7.8 ■ Divorce Rates in Oklahoma Before and After the Oklahoma City Bombing



Source: Nakonezny et al. 2004. "Did Divorces Decline After the Oklahoma City Bombing?" *Journal of Marriage and Family* 66:90–100. Copyright © 2004, John Wiley & Sons. Reprinted with permission.

outcomes, using a regression–discontinuity design. Probationers were assigned to this program if they exceeded a certain value on a recidivism risk score. Jalbert and her colleagues found that at the point for probationers with risk scores near the cutoff score for program participation, recidivism dropped by 25.5% after six months for those who were admitted to the program.

Summary: Causality in Quasi-Experiments

Let's now examine how well quasi-experiments meet the criteria for identifying a nomothetic cause and the two additional challenges to identifying a nomothetic cause (introduced in Chapter 6).

Association Between the Hypothesized Independent and Dependent Variables

Quasi-experiments can provide evidence of association between the independent and dependent variables that is as unambiguous as that provided by a true experiment.

Time Order of Effects of One Variable on the Others

This is a strength of the various quasi-experimental before-and-after designs, but we cannot be as sure of correctly identifying the time order of effects with nonequivalent control group designs because we cannot be certain that some features of the groups did not attract individuals to them who differed at the outset. This is a much greater problem with ex post facto control group designs (see next section).

Nonspurious Relationships Between Variables

We cannot entirely meet this challenge with a quasi-experimental design because we cannot be certain to rule out all potentially extraneous influences with either nonequivalent control group designs or before-and-after designs. Nonetheless, the criteria for these designs do give us considerable confidence that most extraneous influences could not have occurred. Ex post facto control group designs give us much less confidence about the occurrence of extraneous influences

because of the likelihood of self-selection into the groups (see next section). This is why most researchers do not consider ex post facto control group designs to be quasi-experimental.

Mechanism That Creates the Causal Effect

The features of quasi-experiments do not in themselves allow identification of causal mechanisms; however, the repeated measures design does provide a means for testing hypotheses about the causal mechanism.

Context in Which Change Occurs

The quasi-experimental designs that involve multiple groups can provide a great deal of information about the importance of context, as long as the researcher measures contextual variables.

Nonexperimental Ex Post Facto Control Group Designs

It's important now to distinguish a nonexperimental design that is sometimes confused with the quasi-experimental nonequivalent control group design. The **ex post facto control group design** is like the nonequivalent control group design, but it is missing one key feature of quasi-experimental designs. Unlike nonequivalent control group designs, individuals may decide themselves whether to enter the "treatment" or control group. As a result, in ex post facto (after the fact) designs, the people who join the treatment group may differ because of what attracted them to the group initially rather than because of their experience in the group. However, in some studies, we may conclude that the treatment and control groups are so similar at the outset that causal effects can be tested (Rossi and Freeman 1989:343–344).

Cohen and Ledford (1994) studied the effectiveness of self-managing teams in a telecommunications company with an ex post facto design (see Table 7.1). They compared work teams they rated as self-managing with those they found to be traditionally managed (meaning that a manager was responsible for the team's decisions). Cohen and Ledford found the self-reported quality of work-life was higher in the self-managed groups than in the traditionally managed groups.

TABLE 7.1 ■ Ex Post Facto Control Group Design

Self-managing work teams (Cohen and Ledford 1994)			
Experimental group	O1	X	O2
Comparison group	O1		O2
Experimental group	Pretest: Measures of satisfaction and productivity	Self-managing work team	Posttest: Measures of satisfaction and productivity
Comparison group	Pretest: Measures of satisfaction and productivity	Traditionally managed work team	Posttest: Measures of satisfaction and productivity
Key: O = Observation (pretest or posttest) X = Experimental treatment			

Source: Based on Cohen and Ledford (1994).

What distinguishes this study design from a quasi-experimental design is the fact that the teams themselves and their managers had some influence on how they were managed. As the researchers noted, "If the groups which were already high performers were the ones selected to be self-managing teams, then the findings could be due to a selection bias rather than any effects of self-management" (Cohen and Ledford 1994:34). Thus, preexisting characteristics of employees and managers or their team composition might have influenced which "treatment" they received, as well as the outcomes achieved. This leaves us less certain about the effect of the treatment itself.

CAUSAL (INTERNAL) VALIDITY IN EXPERIMENTS

An experiment's ability to yield valid conclusions about causal effects is determined by the comparability of its experimental and comparison groups. First, of course, a comparison group must be created. Second, this comparison group must be so similar to the experimental group or groups that it can show in the posttest what the experimental group would have been like if it had not received the experimental treatment—if the independent variable had not varied. You now know that randomization is used to create a comparison group that is identical to the experimental group at the start of the experiment— with a certain margin of error that occurs with a process of random assignment. For this reason, a true experiment—a design with random assignment—is prone to fewer sources of internal invalidity than a quasi-experiment is.

Several sources of internal invalidity are considerably reduced by a research design that has a comparison group, but others are likely to occur unless a true experimental design with random assignment is used.

Sources of Internal Invalidity Reduced by a Comparison Group

The types of problem that can largely be eliminated by having a comparison group as well as a treatment group are those that arise during the study period itself (Campbell and Stanley 1966:8). Something unanticipated may happen (an effect of "history"), the pretest may have an unanticipated effect on subsequent posttests ("testing"), or the measurement instrument itself may perform differently a second time ("instrumentation"). Also, the participants themselves may change over time ("maturation") or simply decline to their normal levels of performance ("regression effects"). Each of these potential sources of internal invalidity is explained in more detail:

1. *History:* **External events** during the experiment (things that happen outside the experiment) can change subjects' outcome scores. Examples are newsworthy events that have to do with the focus of an experiment and major disasters to which subjects are exposed. This problem is often referred to as a **history effect** —that is, history during the experiment.

   Features of how the treatment is delivered can result in an effect of history, or external events, when there is a comparison, including in true and quasi-experimental designs. For example, in an experiment in which subjects go to a special location for a treatment, something in that location unrelated to the treatment might influence these subjects. In a study of job search services, Price, Van Ryn, and Vinokur's (1992) found that experimental and comparison group subjects differed in whether they attended special seminars, so external events could have happened to subjects in the experimental

group that might not have happened to those in the control group. Perhaps program participants witnessed a robbery outside the seminar building one day, and their orientations changed as a result. External events are a major concern in evaluation studies that compare programs in different cities or states (Hunt 1985:276–277).

2. *Testing:* Taking the pretest can in itself influence posttest scores. Subjects may learn something or be sensitized to an issue by the pretest and, as a result, respond differently the next time they are asked the same questions on the posttest.

3. *Instrumentation:* If the instrument used to measure the dependent variable changes in performance between the pretest and posttest, the result is termed a problem of instrumentation. For example, observers rating the behavior of students in a classroom may grow accustomed to the level of disruptions and therefore rate the same behavior as more appropriate in the posttest than they did in the pretest. This is like the problem of testing, except that it may be possible to reduce the effect of instrumentation by increasing control over the measurement instrument.

4. *Maturation:* Changes in outcome scores during experiments that involve a lengthy treatment period may result from maturation. Subjects may age, gain experience, or grow in knowledge, all as part of a natural maturational experience, and thus respond differently on the posttest than on the pretest.

5. *Regression:* People experience cyclical or episodic changes that result in different posttest scores, a phenomenon known as a **regression effect**. Subjects who are chosen for a study because they received very low scores on a test may show improvement in the posttest, on average, simply because some of the low scorers were having a bad day. Conversely, individuals selected for an experiment because they are suffering from tooth decay will not show improvement in the posttest because a decaying tooth is not likely to improve in the natural course of things. It is hard, in many cases, to know whether a phenomenon is subject to naturally occurring fluctuations, so the possibility of regression effects should be considered whenever subjects are selected because of their extremely high or low values on the outcome variable (Mohr 1992:56, 71–79).

History, testing, maturation, instrumentation, and regression effects could explain any change over time in most of the before-and-after designs because these designs do not have a comparison group. Repeated measures panel studies and time series designs are better in this regard because they allow the researcher to trace the pattern of change or stability in the dependent variable until and after the treatment. However, it is much more desirable to have a comparison group that, like the treatment group, will also be affected by these sources of internal invalidity and so can control for them. Of course, these factors are not a problem in true experiments because those have an experimental group and the comparison group.

Sources of Internal Invalidity Reduced by Randomization

You have already learned the purpose of randomization, or random assignment to the experimental and comparison groups, is to equate the two or more groups at the start of the experiment. The goal is to eliminate the effect of **selection bias**.

6. *Selection bias:* The composition of the experimental and comparison groups in a true experiment is unlikely to be affected by selection bias. Randomization equates the

groups' characteristics, although with some possibility for error due to chance. The likelihood of difference due to chance can be identified with appropriate statistics.

When subjects are not assigned randomly to treatment and comparison groups, as in nonequivalent control group designs, the threat of selection bias is very great. Even if the researcher selects a comparison group that matches the treatment group on important variables, there is no guarantee that the groups were similar initially in the dependent variable or in some other characteristic that ultimately influences posttest scores. However, a pretest helps the researchers determine and control for selection bias. Because most variables that might influence outcome scores will also have influenced scores on the pretest, statistically controlling for the pretest scores also controls many of the unmeasured variables that might have influenced the posttest scores. The potential for selection bias is much greater with an ex post facto control group design because participants have the ability to select the group they enter based on the treatment they expect to receive.

Sources of Internal Invalidity That Require Attention While the Experiment Is in Progress

Even in a research design that involves a comparison group and random assignment, whether or not there is a pretest, the experimental and comparison groups can become different over time because of changes in group membership, interaction between members of the experimental and comparison groups, or effects that are related to the treatment but are not the treatment itself.

7. *Differential attrition:* This problem occurs when the groups become different after the experiment begins because more participants drop out of one of the groups than out of the other(s) for various reasons. **Differential attrition (mortality)** is not a likely problem in a laboratory experiment that occurs in one session, such as Milgram's (1974) study of obedience to authority, but some experiments continue over time. Subjects who experience the experimental condition may become more motivated than comparison subjects are to continue in the experiment and so be less likely to drop out.

You learned in Chapter 6 that attrition can be a major problem in longitudinal designs, whether these are simple panel studies or quasi-experimental repeated measures designs. When many subjects have left a panel between the pretest and posttest (or between any repeated measures in the study), a comparison of the average differences between subjects at the start and end of the study period may mislead us to think the subjects have changed, when what actually happened is that subjects who had dropped out of the study were different from those who remained in it. For example, people with less education and who are not married have been less likely to continue in the large ongoing Panel Study of Income Dynamics (Lillard and Panis 1998:442). Statistical adjustments can reduce the effects of panel attrition, but it is always important to compare the characteristics of study dropouts and those who remain for follow-up.

When the independent variable in an experimental study is a treatment or an exposure to something over time, bias caused by differential attrition can be reduced by using **intent-to-treat analysis**. In this type of analysis, outcomes are compared between all participants who started in the experimental group—those who were intended to receive the treatment—and all participants who started in the control group, whether or not any of these participants left the study before the treatment was fully delivered (Shrout 2011:7). Of course, an intent-to-treat analysis is likely to reduce the researcher's estimate of the treatment effect because some participants will

not have received the full treatment, but it provides a more realistic estimate of effects that are likely to occur if the treatment is administered to persons who can leave before the experiment is over (Bloom 2008:120).

8. *Contamination:* When the comparison group in an experiment is in some way affected by or affects the treatment group, there is a problem with **contamination**. Contamination is not ruled out by the basic features of experimental and quasi-experimental designs, but careful inspection of the research design can determine how much it is likely to be a problem in a particular experiment. This problem basically arises from the failure to control adequately the conditions of the experiment. If the experiment is conducted in a laboratory, if members of the experimental group and the comparison group have no contact while the study is in progress, and if the treatment is relatively brief, contamination is not likely to be a problem. To the degree that these conditions are not met, the likelihood of contamination will increase. For example, contamination was a potential problem in a field-based study by Price et al. (1992) about the effects of a job search training program on the risk of depression among newly unemployed persons. Because the members of both the experimental group (who received the training) and the control group (who did not receive the training) used the same unemployment offices, they could have talked to each other about their experiences while the study was in progress.

9. *Compensatory rivalry:* A problem related to contamination, also termed the **John Henry effect**, can occur when comparison group members are aware they are being denied some advantages and, in response, increase their efforts to compensate for this denial (Cook and Campbell 1979:55).

10. *Demoralization:* This problem involves the opposite reaction to compensatory rivalry; that is, comparison group participants discover they are being denied some treatments they believe are valuable and as a result, they feel demoralized and perform worse than expected. Both **compensatory rivalry** and **demoralization** thus distort the impact of the experimental treatment.

11. *Expectancies of experimental staff:* Change among experimental subjects may result from the positive expectations of the experimental staff who are delivering the treatment rather than from the treatment itself. **Expectancies of experimental staff** may alter the experimental results if staff—even well-trained staff—convey their enthusiasm for an experimental program to the subjects in subtle ways. This is a special concern in evaluation research, when program staff and researchers may be biased in favor of the program for which they work and eager to believe their work is helping clients. Such positive staff expectations thus create a self-fulfilling prophecy. However, in experiments on the effects of treatments such as medical drugs, **double-blind procedures** can be used: Staff delivering the treatments do not know which subjects are getting the treatment and which are receiving a placebo, something that looks like the treatment but has no effect.

12. *Placebo effect:* Treatment misidentification may occur when subjects receive a treatment they consider likely to be beneficial and improve because of that expectation rather than because of the treatment itself. In medical research, where the **placebo effect** often results from a chemically inert substance that looks like the experimental drug

but has no direct physiological effect, some research has indicated the placebo effect itself produces positive health effects in many patients suffering from relatively mild medical problems (Goleman 1993a:C3). It is not clear that these improvements are really any greater than what the patients would have experienced without the placebo (Hrobjartsson and Gotzsche 2001). In any case, it is possible for placebo effects to occur in social science research also, so, when possible, experimental researchers can reduce this threat to internal validity by using multiple comparison groups, some of which do not receive a "placebo" form of the treatment.

13. *Hawthorne effect:* Members of the treatment group may change relative to the dependent variable because their participation in the study makes them feel special. This problem can occur when treatment group members compare their situation with that of members of the control group, who are not receiving the treatment; in this situation, it would be a type of contamination effect. But experimental group members might feel special simply because they are in the experiment. This is termed a **Hawthorne effect**, after a famous productivity experiment at the Hawthorne electric plant outside Chicago. As the story has been told, the workers worked harder no matter what physical or economic conditions the researchers changed to influence productivity; the motivation for the harder work simply seemed to be that the workers felt special because of being in the experiment (Whyt 1955:34).

14. Let me quickly add that a careful review of the actual Hawthorne results shows there wasn't really a clear effect of participating in that experiment (Jone 1992). The Hawthorne effect was itself mostly a matter of misinterpretation and hype. But we can never ignore the possibility that participation in an experiment may itself change participants' orientations and behavior. This is a particular concern in evaluation research when program clients know the research findings may affect the chances for further program funding.

GENERALIZABILITY IN EXPERIMENTS

The need for generalizable findings can be thought of as the Achilles' heel of true experimental design. The design components that are essential for a true experiment and that minimize the threats to causal validity make it more difficult to achieve sample generalizability (being able to apply the findings to some clearly defined larger population) and cross-population generalizability (generalizing across subgroups and to other populations and settings). As a result, findings are often not replicated when experiments are repeated in different settings (Finkel, Eastwick, and Reis 2017)—as you learned in relation to the Sherman and Berk experiments on the police response to domestic violence (see Chapter 2). Nonetheless, no one conducts experiments just to find out how freshman psychology students react to confrontation (or some other experimental "treatment") at your university. Experimental researchers are seeking to learn about general processes, so we have to consider ways to improve the generalizability of their results.

Sample Generalizability

Subjects who can be recruited for a laboratory experiment, randomly assigned to a group, and kept under carefully controlled conditions for the study's duration are unlikely to be a

representative sample of any large population of interest to social scientists. Can they be expected to react to the experimental treatment in the same way as members of the larger population? The generalizability of the treatment and of the setting for the experiment also must be considered (Cook and Campbell 1979:73–74). The more artificial the experimental arrangements are, the greater the problem will be (Campbell and Stanley 1966:20–21).

A researcher can take steps both before and after an experiment to increase a study's generalizability. Conducting a field experiment, such as Sherman and Berk's (1984) study of arrest in actual domestic violence incidents orBraga et al.'s (2018) study of body-worn cameras, is likely to yield more generalizable findings than are laboratory experiments, for which subjects must volunteer. In some field experiments, participants can even be selected randomly from the population of interest, and, thus, the researchers can achieve results generalizable to that population. For example, some studies of the effects of income supports on the work behavior of poor persons have randomly sampled persons within states before randomly assigning them to experimental and comparison groups. When random selection is not feasible, the researchers may be able to increase generalizability by selecting several different experimental sites that offer marked contrasts on key variables (Cook and Campbell 1979:76–77).

Conducting research online is another way to involve people other than undergraduate students in research studies, but bear in mind that Internet users are more likely to be college educated and male than the general population (Hewson, Vogel, and Laurent 2016:79). WEXTOR (http://wextor.org/wextor/en/) provides an online tool for generating experiments both online and in the laboratory (Reips and Neuhaus 2002), while Mechanical Turk provides easy access to people who have signed up to conduct tasks online and can be recruited for online experiments (https://www.mturk.com/mturk/welcome). Extensive pretesting of web-based experiments is critical to identify and resolve problems (Hewson et al. 2016:136–137).

Factorial Surveys

Factorial surveys embed the features of true experiments into a survey design to maximize generalizability. In the most common type of factorial survey, respondents are asked for their likely responses to one or more vignettes about hypothetical situations. The content of these vignettes is varied randomly among survey respondents to create "treatment groups" that differ in particular variables reflected in the vignettes. This is the approach used in the studies of hiring decisions that introduced this chapter.

Van Hoye and Lievens (2003) also used a factorial survey design to test the effect of job applicants' sexual orientation on ratings of their hirability by professionals who make personnel decisions. Van Hoye and Lievens first identified 252 actual selection professionals—people involved daily in personnel selection and recruitment—from consulting firms and company human resource departments. The researchers mailed to each of these professionals a packet with four items: (1) a *letter* inviting their participation in the study, (2) a *job posting* that described a company and a particular job opening in that company, (3) a *candidate profile* that described someone ostensibly seeking that job, and (4) a *response form* on which the selection professionals could rate the candidate's hirability.

The experimental component of the survey was created by varying the candidate profiles. Van Hoye and Lievens created nine different candidate profiles. Each profile used very similar language to describe a candidate's gender (they were all male), age, nationality, family situation, education, professional experience, and personality. However, the family situations were varied to distinguish candidates as heterosexual, homosexual, or "possibly homosexual;" single, and older than 30. Other characteristics were varied to distinguish candidates who were "poor,"

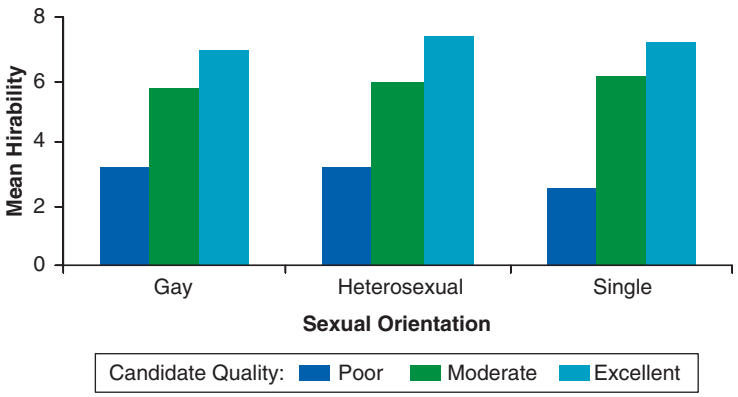
"moderate," and "excellent" matches to the job opening. An example of a profile for a homosexual male who was a "good" candidate for the job included the following language:

1. Personal Data
   Name: Peter Verschaeve
   Gender: Male
   Age: 33 years
   Family situation: Living together with John Vermeulen, fashion designer
2. Educational and Professional Experience
   1990–1993: PUC Diepenbeek—MBA, Marketing Major
   1991now: Human resources manager of an electronics manufacturer
3. Personality
   Peter Verschaeve is self-assured and assertive. He interacts with others in a friendly and warm manner. (Van Hoye and Lievens 2003:27)

The combination of three different descriptions of family situations and three different levels of candidate quality resulted in nine different candidate profiles. Each selection professional was randomly assigned to receive one of these nine candidate profiles. As a result, there was no relationship between who a particular selection professional was and the type of candidate profile they received.

The results of the study appear in Figure 7.9. The average hirability ratings did not differ between candidates who were gay, heterosexual, or single, but hirability increased in direct relation to candidate quality. Van Hoye and Lievens (2003:26) concluded that selection professionals based their evaluations of written candidate profiles on candidate quality, not on their sexual orientation—at least in Flanders, Belgium.

FIGURE 7.9 ■ Average Hirability Ratings in Relation to Candidate Quality and Sexual Orientation



Source: Van Hoye, Greet, and Filip Lievens. 2003. "The Effects of Sexual Orientation on Hirability Ratings: An Experimental Study." *Journal of Business and Psychology* 18:15–30. Copyright © 2003, Human Science Press, Inc. Reprinted with permission from Springer.

Because Van Hoye and Lievens surveyed real selection professionals at their actual workplaces, we can feel more comfortable with the generalizability of their results than if they had just recruited college students for an experiment in a laboratory or MTurkers on the web. However, there is still an important limitation to the generalizability of factorial surveys such as these: A

factorial survey research design indicates only what respondents say they would do in situations that have been described to them. If the selection professionals had to make a recommendation for hiring to an actual employer, we cannot be sure they would act in the same way. So factorial surveys do not completely resolve the problems caused by the difficulty of conducting true experiments with representative samples. Nonetheless, by combining some of the advantages of experimental and survey designs, factorial surveys can provide stronger tests of causal hypotheses than can other surveys and more generalizable findings than can experiments.

External Validity

Researchers are often interested in determining whether treatment effects identified in an experiment hold true for subgroups of subjects and across different populations, times, or settings. Of course, determining that a relationship between the treatment and the outcome variable holds true for certain subgroups does not establish that the relationship also holds true for these subgroups in the larger population, but it suggests the relationship might be externally valid.

We have already seen examples of how the existence of treatment effects in particular subgroups of experimental subjects can help us predict the cross-population generalizability of the findings. For example, Sherman and Berk's (1984) research (see Chapter 2) found that arrest did not deter subsequent domestic violence for unemployed *individuals*; arrest also failed to deter subsequent violence in *communities* with high levels of unemployment. This is a limitation to the generalizability of findings about the impact of arrest.

Interaction of Testing and Treatment

A variant on the problem of external validity occurs when the experimental treatment has an effect only when particular conditions created by the experiment occur. One such problem occurs when the treatment has an effect only if subjects have had the pretest. The pretest sensitizes the subjects to some issue, so when they are exposed to the treatment, they react in a way that differs from how they would have reacted had they not taken the pretest. In other words, testing and treatment interact to produce the outcome. For example, answering questions in a pretest about racial prejudice may sensitize subjects so that when exposed to the experimental treatment, seeing a film about prejudice, their attitudes are different from what they would have been otherwise. In this situation, the treatment truly had an effect, but it would not have had an effect had it been provided without the sensitizing pretest. This possibility can be evaluated with the Solomon four-group design described earlier (see Table 7.2).

TABLE 7.2 ■ Solomon Four-Group Design Testing the Interaction of Pretesting and Treatment

Experimental group:	R	O1	X	O2
Comparison group:	R	O1		O2
Experimental group:	R		X	O2
Comparison group:	R			O2

Key:

R = Random assignment

O = Observation (pretest or posttest)

X = Experimental treatment

As you can see, no single procedure establishes the external validity of experimental results. Ultimately, we must base our evaluation of external validity on the success of replications taking place at different times and places and using different forms of the treatment.

There is always an implicit trade-off in experimental design between maximizing causal validity and generalizability. The more the assignment to treatments is randomized and all experimental conditions are controlled, the less likely it is that the research subjects and setting will be representative of the larger population. College students are easy to recruit and assign to artificial but controlled manipulations, but both practical and ethical concerns preclude this approach with many groups and with respect to many treatments. However, although we need to be skeptical about the generalizability of the results of a single experimental test of a hypothesis, the body of findings accumulated from many experimental tests with different people in different settings can provide a solid basis for generalization (Campbell and Russo 1999:143).

ETHICAL ISSUES IN EXPERIMENTAL RESEARCH

Social science experiments can raise difficult ethical issues. You have already read in Chapter 3 that Philip Zimbardo (2004:34) ended his Stanford prison experiment after only six days, rather than after the planned two weeks, because of the psychological harm that seemed to result from the unexpectedly sadistic behavior of some of the "guards." Although Zimbardo's follow-up research convinced him there had been no lasting harm to subjects, concern about the potential for harm would preclude many such experiments today.

Nonetheless, experimental research continues because of the need for very good evidence about cause–effect relationships to inform social theory as well as social policy. The particular strength of randomized experiments for answering causal questions means they can potentially prevent confusion in social theory and avoid wasting time and resources on ineffective social programs (Mark and Gamble 2009:203).

Two ethical issues are of special importance in experimental research designs. Deception is an essential part of many experimental designs, despite the ethical standard of subjects' informed consents. As a result, contentious debate continues about the interpretation of this standard. In addition, experimental evaluations of social programs pose ethical dilemmas because they require researchers to withhold possibly beneficial treatment from some of the subjects just based on chance (Boruch 1997). In this section, I give special attention to the problems of deception and the distribution of benefits in experimental research.

Deception

Deception is used in social experiments to create more "realistic" treatments, often within the confines of a laboratory. Chapter 3 described Milgram's (1965) use of deception in his classic study of obedience to authority. Volunteers were recruited for what they were told was a study of the learning process, not a study of "obedience to authority." The experimenter told the volunteers they were administering electric shocks to a "student" in the next room, when there were actually neither students nor shocks. Most subjects seemed to believe the deception.

You learned in Chapter 3 that Milgram's description of dehoaxing inflated the consistency with which it was used and the amount of deception that was revealed. You also learned that the dehoaxing increased the negative impact of the experimental experience for some participants. You may therefore be reassured to know that experiments by Barrera and Simpson (2012) found that students' experience of being deceived in a social psychology experiment did not affect their subsequent behavior toward others.

Whether or not you believe you could be deceived in this way, you are not likely to be invited to participate in an experiment such as Milgram's. Current federal regulations preclude deception in research that might trigger such upsetting feelings. However, some deception about a study's purpose is still routine in social psychology laboratories. In many laboratory experiments, the results would be worthless if subjects understood what was really happening to them. Staff making hiring decisions in firms could certainly change their behavior if they knew applications they were reviewing were not from actual job candidates but part of an experiment to test for bias in decision making. The real question is "Is this sufficient justification to allow the use of deception?" Researchers like Silva (2018) turn to MTurk to recruit participants for such studies in part because they don't want to use deceit with professionals actually doing their jobs.

The American Sociological Association's (ASA) *Code of Ethics* (2018) does not discuss experimentation explicitly, but it does highlight the ethical dilemma posed by deceptive research:

> **11.4** Use of Deception in Research
>
> **(a)** Deception can include misleading participants about the research procedures and/or not providing all relevant information about the research. Sociologists do not use deceptive techniques unless they have determined that the following conditions have been met (1) the research involves no more than minimal risk to research participants; (2) deception is justified by the study's prospective scientific, educational, or applied value; (3) equally effective alternative procedures that do not use deception are not feasible; and (4) they have obtained the approval of an authoritative body with expertise on the ethics of social science research such as an institutional review boards.
>
> **(b)** Sociologists do not deceive research participants about significant aspects of the research that would affect their willingness to participate, such as physical risks, discomfort, or unpleasant emotional experiences.
>
> **(c)** When deception is an integral feature of the design and conduct of research, sociologists attempt to correct any misconception that research participants may have as soon as possible, and no later than at the conclusion of the research. This debriefing may be waived if correcting the deception can potentially cause problems for the participants or if debriefing is not possible.
>
> **(d)** Sociologists generally disclose their identities to research participants. On rare occasions, sociologists may need to conceal their identities in order to undertake research that could not practicably be carried out were they to be known as researchers. Under such circumstances, sociologists undertake the research if it involves no more than minimal additional risk for the research participants, and if they have obtained approval from an authoritative body with expertise on the ethics of social science research such as an institutional review board. Under such circumstances, confidentiality must be maintained (p. 14)

Thus, the ASA approach is to allow deception when it is unlikely to cause harm, is necessary for the research, and is followed by an adequate explanation after the experiment is over. As you learned in Chapter 3, the newly revised federal human subjects research standards make an explicit exemption for "research involving benign behavioral interventions in conjunction with the collection of information from an adult subject through verbal or written responses... or audiovisual recording" when the subject had agreed to the intervention (*Federal Register*

2017:7261, 7262, 7264). This would exempt the typical social psychology laboratory experiment from review but not experiments involving interventions like Milgram's or Zimbardo's, which would not be considered "benign."

David Willer and Henry A. Walker (2007, "Debriefing") pay particular attention to debriefing after deception in their book about experimental research. They argue that every experiment involving deception should be followed immediately for each participant with dehoaxing, in which the deception is explained, and then by desensitization, in which all the participants' questions are answered to their satisfaction and those participants who still feel aggrieved are directed to a university authority to file a complaint or to a counselor for help with their feelings. This is sound advice.

Debriefing is a special concern in experiments conducted on the web since participants who withdraw early are usually no longer available. It is therefore good practice to require participants in web experiments (and surveys) to provide an e-mail address initially, so they can be contacted again. The website where the experiment is provided should also be programmed so termination before the experiment ends takes the participant immediately to a debriefing site, where information about the experiment is provided (Toepoel 2016:46).

Selective Distribution of Benefits

Field experiments conducted to evaluate social programs also can involve issues of informed consent (Hunt 1985:275–276). One ethical issue that is somewhat unique to field experiments is the **selective distribution of benefits**: How much are subjects harmed by the way treatments are distributed in the experiment? For example, Sherman and Berk's (1984) experiment and its successors required police to make arrests in domestic violence cases largely based on a random process. When arrests were not made, did the subjects' abused spouses suffer?

Is it ethical to give some potentially advantageous or disadvantageous treatment to people on a random basis? Random distribution of benefits is justified when the researchers do not know whether some treatment actually is beneficial or not—and, of course, it is the goal of the experiment to find out (Mark and Gamble 2009:205). Chance is a more reasonable basis for distributing the treatment than many other bases. Also, if insufficient resources are available to fully fund a benefit for every eligible person, then distribution of the benefit on the basis of chance to equally needy persons is ethically defensible (Boruch 1997:66–67).

The extent to which participation was voluntary varied in the field studies discussed in this chapter. Potential participants in a study by Price et al. (1992) of job search training for unemployed persons signed a detailed consent form in which they agreed to participate in a study involving random assignment to one of the two types of job search help. However, researchers only accepted into the study persons who expressed an equal preference for the job search seminar and the mailed job materials used for the control group.

Thus, Price et al. (1992) avoided the problem of not acceding to subjects' preferences. It, therefore, doesn't seem at all unethical that the researchers gave treatment to only some of the subjects. As it turned out, subjects did benefit from the experimental treatment (the workshops). Now that the study has been conducted, government bodies will have a basis for expecting that tax dollars spent on job search workshops for the unemployed will have a beneficial impact. If this knowledge results in more such programs, the benefit of the experiment will have been considerable, indeed.

Unlike the subjects in the Price et al. (1992) study, individuals who were the subjects of domestic violence complaints in the Sherman and Berk (1984) study had no choice about being arrested or receiving a warning, nor were they aware they were in a research study.

Perhaps it seems unreasonable to let a random procedure determine how police resolve cases of domestic violence. And, indeed, it would be unreasonable if this procedure were a regular police practice. The Sherman and Berk (1984) experiment and its successors do pass ethical muster, however, when seen for what they were: a way of learning how to increase the effectiveness of police responses to this all-too-common crime (Mark and Gamble 2009:205). The initial Sherman and Berk findings encouraged police departments to make many more arrests for these crimes, and the follow-up studies resulted in a better understanding of when arrests are not likely to be effective. The implications of this research may be complex and difficult to implement, but the research provides a much stronger factual basis for policy development.

CONCLUSIONS

True experiments play two critical roles in social science research. First, they are the best research design for testing nomothetic causal hypotheses. Even when conditions preclude use of a true experimental design, many research designs can be improved by adding some experimental components. Second, true experiments also provide a comparison point for evaluating the ability of other research designs to achieve causally valid results.

Despite obvious strengths, true experiments are used infrequently to study many research problems that interest social scientists. There are three basic reasons: (1) The experiments required to test many important hypotheses require far more resources than are often available; (2) many research problems of interest to social scientists are not amenable to experimental designs, for reasons ranging from ethical considerations to the limited possibilities for randomly assigning people to different conditions in the real world; and (3) the requirements of experimental design usually preclude large-scale studies and so limit generalizability to a degree that is unacceptable to many social scientists. Quasi-experiments can be an excellent design alternative.

Although it may be possible to test a hypothesis with an experiment, it may not always be desirable to do so. When a social program is first being developed and its elements are in flux, it is not a good idea to begin a large evaluation study that cannot possibly succeed unless the program design remains constant. Researchers should wait until the program design stabilizes somewhat. It also does not make sense for evaluation researchers to test the impact of programs that cannot actually be implemented or that are unlikely to be implemented in the real world because of financial or political problems (Rossi and Freeman 1989:304–307).

Even laboratory experiments are inadvisable when they do not test the real hypothesis of interest but test instead a limited version amenable to laboratory manipulation. The intersecting complexity of societies, social relationships, and social beings—of people and the groups to which they belong—is so great that it often defies reduction to the simplicity of a laboratory or restriction to the requirements of experimental design. Yet the virtues of experimental designs mean they should always be considered when explanatory research is planned.

KEY TERMS

Assignment variable (p. 241)
Before-and-after design
Comparison group (p. 230)
Compensatory rivalry (John Henry effect)
Contamination (p. 247)
Control group (p. 230)
Demoralization (p. 247)
Differential attrition (mortality) (p. 246)

Double-blind procedure
Ex post facto control group design (p. 243)
Expectancies of experimental staff (p. 247)
Experimental group (p. 230)
External events (p. 244)
Factorial survey
Field experiment
Hawthorne effect (p. 248)
History effect (p. 244)
Intent-to-treat analysis (p. 246)
Matching (p. 233)
Multiple group before-and-after design (p. 240)
Nonequivalent control group design
Placebo effect (p. 247)
Posttest (p. 232)
Pretest (p. 232)
Quasi-experimental design (p. 238)
Randomized comparative change design (p. 232)
Randomized comparative posttest design (p. 232)
Regression effect (p. 245)
Regression–discontinuity design (p. 241)
Repeated measures panel design
Selection bias (p. 245)
Selective distribution of benefits (p. 254)
Solomon four-group design (p. 234)
Time series design
True experiment

HIGHLIGHTS

LO 7.1 Identify the essential components of a true experimental design.

- Experimental research designs have three essential components: (1) use of at least two groups of subjects for comparison, (2) measurement of the change that occurs as a result of the experimental treatment, and (3) use of random assignment.
- True experiments must have at least one experimental group (subjects who receive some treatment) and at least one comparison group (subjects with whom the experimental group can be compared). The comparison group differs from the experimental group in one or more independent variables, whose effects are being tested.
- All true experiments have a posttest—that is, measurement of the outcome in both groups after the experimental group has received the treatment. Many true experiments also have pretests that measure the dependent variable before the experimental intervention.
- A randomized comparison group can provide a good estimate of the counterfactual—the outcome that would have occurred if the subjects who were exposed to the treatment actually had not been exposed but otherwise had had the same experiences (Mohr 1992; Rossi and Freeman 1989:229).
- In addition, experiments may include identification of a causal mechanism and control over experimental conditions.

LO 7.2 Assess the advantages and disadvantages of using a quasi-experimental design.

- Quasi-experiments are an option when testing a hypothesis with a true experimental design is not feasible with the desired participants and in the desired setting.
- Quasi-experiments can have a nonequivalent control group design in which a comparison group is selected to be as comparable as possible to the treatment group. Two selection methods can be used: aggregate matching and individual matching.
  - Aggregate matching creates a comparison group that matches the treatment group in many ways but differs in exposure to the treatment.

    - Individual matching creates a comparison group by matching individual cases in the treatment group with similar individuals in the comparison group.
- Quasi-experiments can have a multiple group before-and-after design in which several before-and-after comparisons are made involving the same independent and dependent variables but different groups.
- Advantages
    - Quasi-experiments include features that maximize the comparability of the control and experimental groups.
    - Quasi-experiments include features that make it unlikely that self-selection determines group membership.
    - Quasi-experiments may provide more generalizable results than true experiments.
    - Quasi-experiments can provide evidence of association between the independent and dependent variables that is as unambiguous as that provided by a true experiment.
    - Before-and-after designs can determine the time order of effects of one variable on the others.
- Disadvantages
    - Quasi-experiments are more prone to some problems of internal validity because of their lack of random assignment (although some quasi-experimental designs allow the researcher to rule out almost as many potential sources of internal validity as does a true experiment).
    - We cannot entirely ensure nonspurious relationships between variables with a quasi-experimental design because we cannot be certain to rule out all potentially extraneous influences.
    - The features of quasi-experiments do not in themselves allow identification of causal mechanisms.
    - In nonequivalent control group designs, we cannot be sure of correctly identifying the time order of effects because we cannot be certain that some features of the groups did not attract individuals to them who differed at the outset.

LO 7.3 State the features of research design that can reduce threats to internal validity.

- A true experiment—a design with random assignment—is prone to fewer sources of internal invalidity than a quasi-experiment is.
- Threats to internal invalidity can be reduced by having a comparison group as well as a treatment group during the study period.
- Threats to internal validity can be reduced by randomization. This reduces the effect of selection bias. Selection bias occurs when characteristics of experimental and comparison group subjects differ in any way that influences the outcome.
- Threats to internal validity can be reduced by monitoring if the experimental and comparison groups become different over time because of changes in group membership, interaction between members of the experimental and comparison groups, or effects that are related to the treatment but are not the treatment itself.

LO 7.4 Explain why generalizability can be such a challenge in experimental research.

- The design components that are essential for a true experiment and that minimize the threats to causal validity make it more difficult to achieve sample generalizability (being able to apply the findings to some clearly defined larger

population) and cross-population generalizability (generalizing across subgroups and to other populations and settings).

- Subjects who can be recruited for a laboratory experiment, randomly assigned to a group, and kept under carefully controlled conditions for the study's duration are unlikely to be a representative sample of any large population of interest to social scientists.
- In terms of external validity, determining that a relationship between the treatment and the outcome variable holds true for certain subgroups does not establish that the relationship also holds true for these subgroups in the larger population, but it suggests the relationship might be externally valid across these subgroups.
- A variant on the problem of external validity occurs when the experimental treatment has an effect only when particular conditions created by the experiment occur. One such problem occurs when the treatment has an effect only if subjects have had the pretest. This problem can be avoided by using the Solomon Four-Group Design.

LO 7.5 Critique the most distinctive ethical challenges in experimental research.

- Subject deception is common in laboratory experiments and poses unique ethical issues. The ASA approach is to allow deception when it is unlikely to cause harm, is necessary for the research, and is followed by an adequate explanation after the experiment is over.
- In field experiments, a common ethical problem is the selective distribution of benefits. Subjects can be harmed by the way treatments are distributed in an experiment. Random assignment may be the fairest way of allocating treatment when treatment openings are insufficient for all eligible individuals and when the efficacy of the treatment is unknown.

DISCUSSION QUESTIONS

1. You probably know that some people refused COVID-19 vaccines because they said they did not want to be used as "guinea pigs" by researchers. Some explicitly explained this attitude as reflecting the way that participants in the Tuskegee research were used. Yet thousands of people around the world participated in COVID-19 vaccine trials, and it is only because they did so that health care leaders were able to determine which vaccines were safe and effective. What strategies would you suggest for reducing fear of participating in research sponsored by organizations and researchers with appropriate expertise and oversight? Which ethical standards would you suggest emphasizing in recruiting participants?

2. Should social scientists be permitted to conduct replications of Milgram's obedience experiments? Zimbardo's prison simulation? Can you justify such research as permissible within the current ASA ethical standards? If not, do you believe these standards should be altered to permit Milgram-type research?

3. How do you evaluate the current ASA ethical code? Is it too strict or too lenient or just about right? Are the enforcement provisions adequate? What provisions could be strengthened?

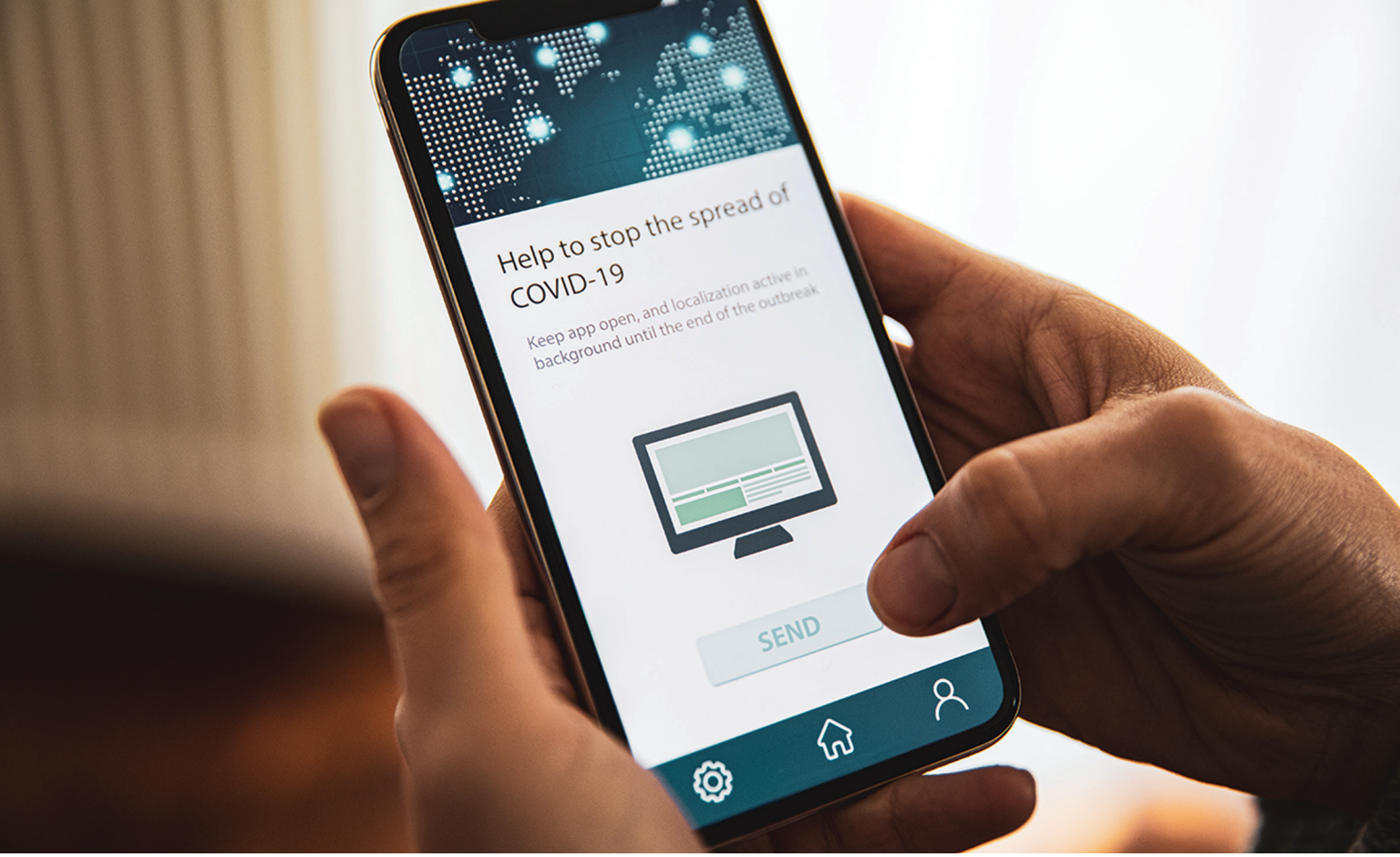
4. Why does unethical research occur? Is it inherent in science? Does it reflect "human nature"? What makes ethical research more or less likely?

5. Does debriefing solve the problem of subject deception? How much must researchers reveal after the experiment is over as well as before it begins?

SPSS EXERCISES

1. Consider three variables in the GSS2020 survey: helpblk_2, income16_2, and owngun_2. Review the corresponding variable labels in the "Variable View" in the GSS2020 file you are using so that you know what questions were used to measure these variables. Are there any ethical issues involved in asking these questions? Do you imagine that some respondents would be more likely to give untruthful answers to these questions? Explain your answers.

2. Examine the cross-tabulation between fundamentalist beliefs and support for the right to abortion after a rape and support for allowing an antireligionist to speak (Analyze→Descriptive Statistics→Crosstabs→ (Rows = abrape_2, spkath_2; Columns = polviews_2; Cells/Percentages = columns). Summarize what you have learned about the association between political conservatism and feelings about abortion and allowing an antireligionist to speak.


Help to stop the spread of COVID-19
Keep app open, and localization active in background until the end of the outbreak
SEND

©iStockphot.com/LeoPatrizi

8 SURVEY RESEARCH

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

8.1 Summarize the reasons for the popularity of survey research

8.2 Report the four types of errors of observation that must be minimized in a successful survey

8.3 Describe the basic principles for developing clear and meaningful survey questions

8.4 Defend the rationale for using multi-question indexes

8.5 Appraise the best practices to reduce error through careful survey design

8.6 Compare the strengths and weaknesses of the five basic social science survey designs

8.7 Discuss the key ethical issues in survey research

"How concerned are you, if at all, about the coronavirus or COVID-19 outbreak?" Sixty-six percent of Americans responding to a survey in early April 2020 reported they were "extremely" or "very" concerned. In the same survey, 81% rated the risk of "attending in-person gatherings of friends and family outside your household" as "large" or "moderate." By late October 2021, the percentage reporting being extremely or very concerned with the coronavirus had declined to 39%, and only 45% rated the risk of in-person gatherings as large or moderate (Jackson 2021).

Many survey projects in the United States and around the globe charted the impact of the COVID-19 pandemic. From concerns about infection to behaviors shaping risk, from economic impact to mental health consequences, survey research allowed social scientists to understand the pandemic's social patterning and to inform policies to mitigate its adverse effects. I begin this chapter by highlighting such reasons for using survey methods. I then illustrate some key features of survey research with a survey about the pandemic's effects. I then present guidelines for writing survey questions, explain the major steps in questionnaire design, and discuss the distinctive characteristics of five types of surveys—highlighting the unique problems attending each one and suggesting some possible solutions. I give particular attention to the ways in which smartphones and the Internet and other technological advances have changed survey research in recent decades. I discuss related ethics issues in the final section. By the chapter's end, you will be well on your way to becoming an informed consumer of survey reports and a knowledgeable developer of survey designs. As you read the chapter, I also hope you will reflect on how survey research helped shape responses to the pandemic.

SURVEY RESEARCH IN THE SOCIAL SCIENCES

Survey research involves the collection of information from a sample of individuals through their responses to questions. As you probably have observed, survey research is a multibillion-dollar industry that shapes what we read in the newspapers, see on TV, and find in government reports (Converse 1984; Tourangeau 2004:776).

Attractions of Survey Research

Survey research owes its popularity to three features: versatility, efficiency, and generalizability. Each of these features is changing as a result of new technologies.

Versatility

First, survey methods are versatile. A well-designed survey can enhance understanding of just about any social issue. Politicians campaigning for election use surveys, as do businesses marketing a product, governments assessing community needs, agencies monitoring program effectiveness, and lawyers seeking to buttress claims of discrimination. You can see at a glance the range of topics and research questions that have been the focus of surveys by spending a few minutes on the websites of major survey organizations like Gallup (www.gallup.com), the National Opinion Research Center at the University of Chicago (www.norc.org), and the Pew Research Center (www.pewresearch.org). The Societal Experts Action Network (SEAN) COVID-19 Survey Archive (https://covid-19.parc.us.com/client/index.html#/) includes links to most surveys about the pandemic.

Efficiency

Surveys also are popular because data can be collected from many people at relatively low cost and, depending on the survey design, relatively quickly. In a two-part survey of Boston residents during the pandemic, funded by the National Science Foundation, my collaborators and I used both mailed paper-and-pencil and web-based survey options, with a cost of about $75 per respondent per completed survey. Many variables can be measured without substantially increasing the time or cost. The 2021 General Social Survey (GSS) had three versions in English and Spanish that measured more than 1,400 variables for 1,620 persons completing the survey on the web (during the pandemic) between December 2020 and May 2021 (NORC 2021).

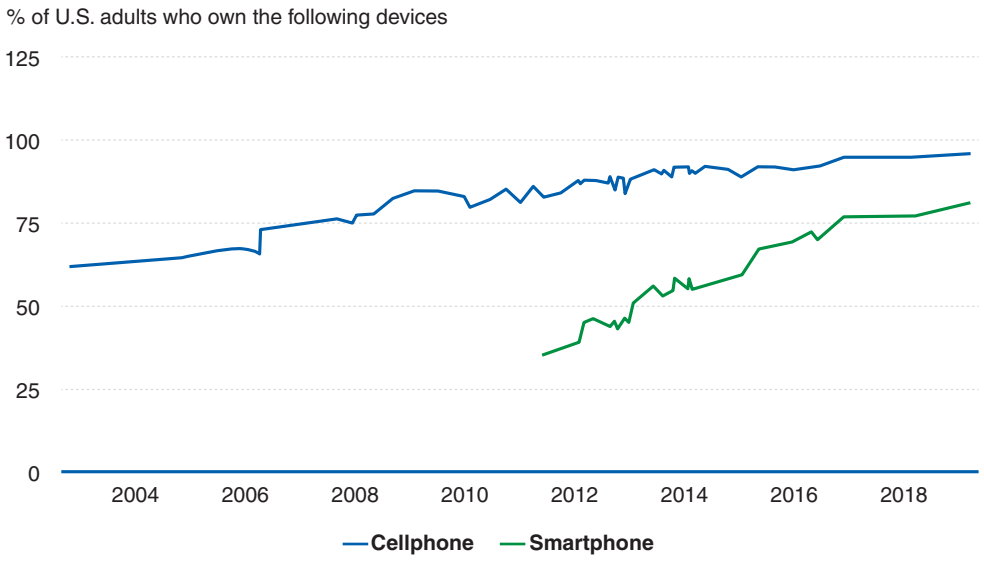
Of course, these efficiencies can be attained only in a place with a reliable communications infrastructure (Labaw 1980:xiii–xiv). A reliable postal service, which is required for mail surveys, generally has been available in the United States—although delays have been a major problem in some low-income communities (Office of Inspector General 2017; and see Lyall 2004:Ar4). Also important to efficiency are the many survey research organizations—about 120 academic and nonprofit organizations in the United States—that provide trained staff and proper equipment (Survey Research Laboratory 2008).

Modern information technology has been a mixed blessing for survey efficiency. The Internet makes it easier to survey some populations, but it leaves out important segments. Caller ID and answering machines make it easy to screen out unwanted calls, but these tools also make it harder to reach people in phone surveys. In addition, although almost all U.S. adults own a cell phone (97% in February 2021)—a large majority (85%) use a smartphone also providing Internet access (Pew Research Center 2021)—and a growing number do not have a landline and so are not listed in a phone directory (Christian et al. 2010; McGeeney and Keeter 2014; U.S. Census Bureau 2013) (see Figure 8.1). As a result of these changes, survey researchers must spend more time and money to reach potential respondents (Tourangeau 2004:781–782).

Generalizability

Survey methods lend themselves to probability sampling from large populations. Thus, survey research is appealing when sample generalizability is a central research goal. Surveys are also useful when cross-population generalizability is a key concern because they allow a range of social

FIGURE 8.1 ■ Percentage of U.S. Adults With Cellphones and Smartphones



Source: Mobile Fact Sheet. Surveys conducted 2002–2019. Data source: Pew Research Center. April 7, 2021. https://www.pewresearch.org/internet/fact-sheet/mobile/.

contexts and subgroups to be sampled. Unfortunately (for survey researchers), different types of surveys are not equally accessible to different population segments. In the United States, households without Internet connections are more likely to be elderly, poor, rural, and to have only a high school education (Anderson and Perrin 2016; de Leeuw 2008:321; U.S. Census Bureau 2013). U.S. households that rely only on landlines tend to be elderly (Lavrakas et al. 2017). As a result, Web surveys and those that rely on phone calls to landline numbers are not likely to be representative of the general population. The foreign-born population in the United States (13% in 2014) also creates a challenge to generalizability, since foreign-language versions of survey forms are often required, thus resulting in unrepresentative samples or increased costs for translating the surveys (Grieco et al. 2012:2; Tourangeau 2004:783; U.S. Census Bureau 2014).

The Omnibus Survey

An omnibus survey shows just how versatile, efficient, and generalizable a survey can be. An **omnibus survey** covers a range of topics of interest to different social scientists, in contrast to the typical survey that is directed at a specific research question. The omnibus survey has multiple sponsors or is designed to generate data useful to a broad segment of the social science community rather than to answer a particular research question. It is usually directed to a sample of some general population, so the questions, about a range of different issues, are appropriate to at least some sample members.

The General Social Survey (GSS) of the National Opinion Research Center at the University of Chicago is one of sociology's most successful omnibus surveys. It is an extensive interview administered biennially to a probability sample of at least 2,000 Americans, with a wide range of questions and topic areas chosen by a board of overseers. Some questions are asked of only a randomly selected subset of respondents. This **split-ballot design** allows more questions without increasing the survey's cost. It also facilitates experiments on the effect of question wording: Different forms of the same question are included in the split-ballot subsets. The GSS is widely

available to universities, instructors, and students (Davis and Smith 1992; NORC 2011). Many other survey data sets are archived by the Inter-university Consortium for Political and Social Research (ICPSR) (more details about the ICPSR are in Chapter 14).

ERRORS IN SURVEY RESEARCH

For a survey to succeed, it must minimize four types of error (Groves 1989:vi, 10–12): (1) poor measurement, (2) nonresponse, (3) inadequate coverage of the population, and (4) sampling error.

Poor Measurement

Measurement error was a key concern in Chapter 4, but there is much more to be learned about how to minimize these errors of observation in the survey process. The theory of satisficing can help us understand the problem. It takes effort to answer survey questions carefully: Respondents must figure out what each question means, then recall relevant information, and finally decide which answer is most appropriate. Survey respondents satisfice when they reduce the effort required to answer a question by interpreting questions superficially and giving what they think will be an acceptable answer (Krosnick 1999:47–548). A tendency to choose responses appearing earlier in a list of responses—a "primacy effect"—is a similar problem (Toepoel 2016:23). Errors in measurement also arise when respondents are unwilling to disclose their feelings and behaviors, unable to remember past events, and misunderstand survey questions. What people say they can do—such as ability to carry out various tasks—is not necessarily consistent with what they are able to do (Schutt 2011b:88). What people report they have done is not necessarily what they have done (Brenner 2012). A natural desire to say "what the interviewer wants to hear" can generate an "acquiescent response bias" among some respondents while others may answer questions about sensitive issues in a way they believe is more socially desirable (Toepoel 2016:23).

Presenting clear and interesting questions in a well-organized questionnaire will help reduce measurement error by encouraging respondents to answer questions carefully and to take seriously the request to participate in the survey. Tailoring questions to the specific population surveyed is also important, so understanding the culture of the population and setting being surveyed is essential (Holbrook et al. 2003; Narayan and Krosnick 1996; Robinson and Leonard 2019). The next section focuses on how to write good survey questions.

Nonresponse

Nonresponse is a major and growing problem in survey research, although it is a problem that varies between survey designs. The perceived benefits of survey participation have declined with decreasing levels of civic engagement, longer work hours, and increasing numbers of survey invitations distributed by email (Groves et al. 2000; Krosnick 1999:539–540; Robinson and Leonard 2019; Tourangeau 2004:782). Perceived costs have increased with the widespread use of telemarketing and the ability of many people to screen out calls from unknown parties with answering machines and caller ID. We will review more specifics about nonresponse in this chapter's sections on different survey methods.

Inadequate Coverage of the Population

A poor sampling frame can invalidate the results of an otherwise well-designed survey. We considered the importance of a good sampling frame in Chapter 5, so I will only discuss special coverage problems of some survey methods in this chapter.

Sampling Error

The process of random sampling can result in differences between the characteristics of the sample members and the population simply based on chance. I introduced this as a topic in Chapter 5. You will learn how to calculate sampling error in Chapter 9.

Although surveys provide an efficient means for investigating a wide range of issues in large and diverse populations, the data they provide are necessarily influenced by these four sources of error. Survey researchers must make every effort to minimize each one.

WRITING SURVEY QUESTIONS

Questions are the centerpiece of survey research. Because the way they are worded can have a great effect on the way they are answered, selecting good questions is the single most important concern for survey researchers. All hope for achieving measurement validity is lost unless the questions in a survey are clear and convey the intended meaning to respondents.

Question-writing should not begin until the purpose for conducting a survey is identified. That purpose should frame the evaluation of all possible questions (Robinson and Leonard 2019). Writing questions for a particular survey might then begin with a brainstorming session or a review of previous surveys. Although most professionally prepared surveys contain previously used questions as well as some new ones, every question considered for inclusion must be reviewed carefully for its clarity and ability to convey the intended meaning. Questions that were clear and meaningful to one population may not be so to another.

Adherence to a few basic principles will go a long way toward ensuring clear and meaningful questions (Alwin and Krosnick 1991).

Avoid Confusing Phrasing

What's a confusing question? Try this one that I received years ago from the Planetary Society in its National Priorities Survey for the U.S. Space Program:

> The Moon may be a place for an eventual scientific base, and even for engineering resources. Setting up a base or mining experiment will cost tens of billions of dollars in the next century. Should the United States pursue further manned and unmanned scientific research projects on the surface of the Moon?
>
> ■ Yes ■ No ■ No opinion

Does a "yes" response mean you favor spending tens of billions of dollars for a base or mining experiment? Does "the next century" refer to the 21st century or to the 100 years after the survey (which was distributed in the 1980s)? Could you favor further research projects on the Moon but oppose funding a scientific base or engineering resources? Are engineering resources supposed to have something to do with a mining experiment? Does a mining experiment occur "on the surface of the Moon"? How do you answer if you favor unmanned scientific research projects on the Moon but not manned projects?

There are several ways to avoid such confusing phrasing. In most cases, a simple direct approach to asking a question minimizes confusion. Use shorter rather than longer words and sentences: "brave" rather than "courageous"; "job concerns" rather than "work-related employment issues" (Dillman 2000:52). Try to keep the total number of words to 20 or fewer and the number of commas to three or fewer (Peterson 2000:50). However, questions should not be

abbreviated in a way that results in confusion: To ask, "In what city or town do you live?" is to focus attention clearly on a specific geographic unit, a specific time, and a specific person (you); by contrast, the simple format,

Residential location: _____, does not do this.

A sure way to muddy the meaning of a question is to use **double negatives**: "Do you *disagree* that there should *not* be a tax increase?" Respondents have a hard time figuring out which response matches their sentiments. For instance, in a survey commissioned by the American Jewish Committee, the Roper polling organization wrote a question about the Holocaust that was carefully worded to be neutral and value free: "Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?" Among a representative sample of adult Americans, 22% answered that it was possible the extermination never happened (Kifner 1994:A12). However, when the Gallup organization rephrased the question to avoid the double negative, giving a brief definition of the Holocaust and then asking, "Do you doubt that the Holocaust actually happened or not?" Only 9% responded they doubted it happened.

So-called **double-barreled questions** are also guaranteed to produce uninterpretable results because they ask two questions but allow only one answer. For example, when a Gallup poll during the Watergate scandal asked the question, "Do you think President Nixon should be impeached and compelled to leave the presidency, or not?" only about a third of Americans supported impeaching President Richard M. Nixon. However, when the question was rephrased to ask respondents if they "think there is enough evidence of possible wrongdoing in the case of President Nixon to bring him to trial before the Senate, or not," over half answered yes. Apparently, the first, double-barreled version of the question confused support for impeaching Nixon—putting him on trial before the Senate—with concluding he was guilty before he had had a chance to defend himself (Kagay and Elder 1992:E5).

It is also important to be sure a question is asked of only the respondents who have the information of experiences needed to answer. For example, before we asked respondents to our survey in the early months of the pandemic about their experiences when they left home, we first asked whether they ever left home (see Figure 8.2). Such **filter questions** create **skip patterns**. Respondents who answer *no* to one question are directed to skip ahead to another question, but

FIGURE 8.2 ■ Filter Questions and Skip Patterns

Thinking about April 2020

Please think back to how things were for you in April, when things were shut down because of the COVID-19 coronavirus.

4. In a typical week in April, how many days did you leave your home for any reason?
   - ☐ All 7 days
   - ☐ 5 or 6 days
   - ☐ 3 or 4 days
   - ☐ 1 or 2 days
   - ☐ None, I never left home → **Go to #8 on Page 2**

5. In a typical week in April, when you left your home, how often did you wear a face mask or covering?
   - ☐ All of the time
   - ☐ Most of the time
   - ☐ About half of the time
   - ☐ Less than half of the time
   - ☐ None of the time

Source: Living in Boston During COVID-19 Survey. Data source: Center for Survey Research, University of Massachusetts Boston. June 2021.

respondents who answer *yes* go on to the **contingent question**. Skip patterns should be indicated clearly with an arrow or other mark in the questionnaire.

Minimize the Risk of Bias

Specific words in survey questions should not trigger biases unless that is the researcher's conscious intent. Biased or loaded words and phrases tend to produce misleading answers. For example, a 1974 survey found that 18% of respondents supported sending U.S. troops "if a situation like Vietnam were to develop in another part of the world." But when the question was reworded to mention sending troops to "stop a communist takeover"—"communist takeover" being a loaded phrase—favorable responses rose to 33% (Schuman and Presser 1981:285).

Responses can also be biased when response alternatives do not reflect the full range of possible sentiment on an issue. When people pick a response choice, they seem to be influenced by where they are placing themselves relative to the other response choices (Robinson and Leonard 2019). For example, the Detroit Area Study (Turner and Martin 1984:252) asked the following question: "People feel differently about making changes in the way our country is run. In order to keep America great, which of these statements do you think is best?" When the only response choices were "We should be very cautious of making changes" and "We should be free to make changes," only 37% said we should be free to make changes. However, when a response choice was added that suggested we should "constantly" make changes, 24% picked that response and another 32% chose the "free to make changes" response, for a total of 56% who seemed open to making changes in the way our country is run (Turner and Martin 1984:252). Including the more extreme positive alternative ("constantly" make changes) made the less extreme positive alternative ("be free to make changes") more attractive.

If the response alternatives for a question fall on a continuum from positive to negative, the number of positive and negative categories should be balanced so that one end of the continuum doesn't seem more attractive than the other (Dillman 2000:57–58). For the same reason, it is also better to state both sides of attitude scales in the question itself: "How satisfied or dissatisfied are you with the intramural sports program here?" (Toepoel 2016:27), with answer choices ranging from "completely satisfied" to "completely dissatisfied." This is called a *bipolar* scale.

Of course, the advice to minimize the risk of bias is intentionally ignored by those who conduct surveys to elicit bias. This is the goal of *push polling*, a technique that has been used in some political campaigns. In a push poll, the pollsters for a candidate call potential voters and ask them a series of questions that convey negative information about the opposing candidate. It's not an actual survey at all—just a propaganda effort—but it casts reputable survey research (and ethical-political polling firms) in a bad light (Connolly and Manning 2001).

Maximize the Utility of Response Categories

Response choices help respondents to understand what the question is about and what types of responses are viewed as relevant (Clark and Schober 1994). Questions with fixed response choices must provide one and only one possible response for everyone who is asked the question—that is, the response choices must be exhaustive and mutually exclusive. Ranges of ages, incomes, years of schooling, and so forth should not overlap and should provide a response option for all respondents.

There are two exceptions to this principle: (1) Filter questions may tell some respondents to skip over a question (the response choices do not have to be exhaustive), and (2) respondents may be asked to "check all that apply" (the response choices are not mutually exclusive). Even these exceptions should be kept to a minimum (Dillman 2000: 63).

Vagueness in the response choices is also to be avoided. Questions about thoughts and feelings will be more reliable if they refer to specific times or events (Turner and Martin 1984:300). Usually a question like "On how many days did you read the newspaper in the last week?" produces more reliable answers than one like "How often do you read the newspaper?" in which response choices of "frequently," "sometimes," and "never" are given (Toepoel 2016:27). Our pandemic survey asked about a series of activities in terms of numbers of days rather than with categories like "frequently" or "sometimes" (see Figure 8.3).

FIGURE 8.3 ■ Questions Measuring Frequency of Activities During the Pandemic

In the last 7 days, how many days did you...	All 7 days ↓	5 or 6 days ↓	3 or 4 days ↓	1 or 2 days ↓	None ↓
ride on a bus, trolley, or subway train	☐	☐	☐	☐	☐
go out to eat at a restaurant, bar, or club	☐	☐	☐	☐	☐
go to a food pantry or food bank for groceries	☐	☐	☐	☐	☐
go to a store for groceries	☐	☐	☐	☐	☐
order food delivered to your home from a restaurant	☐	☐	☐	☐	☐
order food delivered to your home from a grocery store	☐	☐	☐	☐	☐
visit inside someone else's home	☐	☐	☐	☐	☐
go outside for exercise or walking	☐	☐	☐	☐	☐
attend any kind of event where more than 10 people were gathered	☐	☐	☐	☐	☐
have people who do not live with you in your home, either to work or visit	☐	☐	☐	☐	☐

Source: Living in Boston During COVID-19 Survey. Data source: Center for Survey Research, University of Massachusetts Boston. June 2021.

Sometimes, problems with response choices can be corrected by adding questions. For example, if you ask, "How many years of schooling have you completed?" someone who dropped out of high school but completed the requirements for a General Educational Development Test (GED) might not be sure how to respond. By asking a second question, "What is the highest degree you have received?" you can provide the correct alternative for those with a GED as well as for those who graduated from high school.

Adding questions may also improve memory about specific past events. Imagine the problem you might have responding to the question: "How often did you receive help from classmates while preparing for exams or completing assignments during the last month? (very often, somewhat often, occasionally, rarely, or never)." Now, imagine a series of questions that asks you to identify the exams and assignments you had in the past month and, for each one, inquire whether you received each of several types of help from classmates: study suggestions, study sessions, related examples, general encouragement, and so on. The more specific focus on specific exams and assignments should result in more complete recall (Dykema and Schaeffer 2000).

Response choices should be matched to the question they follow and reflect meaningful distinctions. If the question is "How satisfied are you with your job?" the response choices should focus on distinctions between levels of satisfaction and might range from "very satisfied" to "somewhat," "not very," and "not at all satisfied." If one response choice is "somewhat satisfied,"

"moderately satisfied" would not be a good additional response because it does not reflect a meaningful distinction.

One common approach for measures of attitude intensity is to present a statement and then ask respondents to indicate their degree of agreement or disagreement. This approach was popularized by social psychologist Rensis Likert (1932) and is known as a **Likert item**. A Likert item phrases an attitude in terms of one end of a continuum, so the responses ranging from "strongly agree" to "strongly disagree" (or "strongly approve" to "strongly disapprove" [Likert's original terms]) cover the full range of possible agreement. For example, the following statement and questions began a series of Likert items in a questionnaire I used years ago in a study of staff in a homeless shelter.

I. Different people have different opinions about homeless persons and services for them. How much do you agree or disagree with each of the following statements about the persons who use your shelter?

1. Few of the shelter's guests are ready to live on their own.

STRONGLY AGREE.....1

AGREE.....2

NEITHER AGREE NOR DISAGREE.....3

DISAGREE.....4

STRONGLY DISAGREE.....5

...

3. The shelter's guests are mostly people who have had some bad breaks.

STRONGLY AGREE.....1

AGREE.....2

NEITHER AGREE NOR DISAGREE.....3

DISAGREE.....4

STRONGLY DISAGREE.....5

However, responses to Likert-style items can be misleading because of the tendency of people to want to be agreeable (see the next section) and because the meaning of the response choices depend on how the preceding statement is worded (what if I had said "Many" instead of "Few" in question 1, or "sometimes" instead of "mostly" in question 3?).

As a result of these considerations, Likert items should be avoided (Saris 2010) and other words should be used to distinguish response choices on an ordinal scale of attitude intensity (see Figure 8.4). One important decision is whether to use unipolar distinctions, such as "not at all" to "extremely," or bipolar distinctions, such as "very comfortable" to "very uncomfortable." The advantages of using **bipolar response options** are discussed in the next section. How many response categories are desirable? Five categories work well for unipolar ratings, and seven will capture most variations on bipolar ratings (Krosnick 2006; Schaeffer and Presser 2003:78–79). Responses are more reliable when these categories are labeled (**labeled unipolar response options**) rather than identified only by numbers (**unlabeled unipolar response options**) (Krosnick 1999:544; Schaeffer and Presser 2003:78).

Avoid Making Either Disagreement or Agreement Disagreeable

People often tend to "agree" with a statement just to avoid seeming disagreeable. This is termed *agreement bias*, **social desirability bias**, or an *acquiescence effect.* You can see the impact of this

FIGURE 8.4 ■ Labeled Unipolar, Unlabeled Unipolar, and Bipolar Response Options

Original

>Q72a< How free do you feel to disagree with the person who supervises your work? Are you . . .

<1> Not at all free,
<2> Somewhat free,
<3> Largely but not completely free, or
<4> Completely free to disagree?
<7> NO CODED RESPONSE APPLICABLE
<8> DON'T KNOW
<9> REFUSED

Labeled unipolar version

How comfortable do you feel disagreeing with the person who supervises your work? (Please circle one number to indicate your response.)

1. Extremely comfortable
2. Very comfortable
3. Quite comfortable
4. Somewhat comfortable
5. Not at all comfortable

Labeled bipolar version

Do you feel comfortable or uncomfortable disagreeing with the person who supervises your work? (Please circle one number to indicate your response.)

1. Very comfortable
2. Mostly comfortable
3. Slightly comfortable
4. Feel neither comfortable nor uncomfortable
5. Slightly uncomfortable
6. Mostly uncomfortable
7. Very uncomfortable

Unlabeled unipolar version

Please circle a number from 1 to 10 to indicate how comfortable you feel disagreeing with the person who supervises your work. 1 means "not at all comfortable" and 10 means "extremely comfortable."

How comfortable do you feel disagreeing with the person who supervises your work?

Not at all									Extremely
1	2	3	4	5	6	7	8	9	10

Source: Based on Mirowsky and Ross (2001:9).

human tendency in a 1974 University of Michigan Survey Research Center survey that asked who was to blame for crime and lawlessness in the United States (Schuman and Presser 1981:208). When one question stated that individuals were more to blame than social conditions, 60% of the respondents agreed. But when the question was rephrased and respondents were asked, in a balanced fashion, whether individuals or social conditions were more to blame, only 46% chose individuals. Numerous studies of agreement bias suggest that about 10% of respondents will "agree" just to be agreeable, without regard to what they really think (Krosnick 1999:553).

You can take several steps to reduce the likelihood of agreement bias. You should present both sides of attitude scales in the question itself (Dillman 2000:61–62): "In general, do you believe that *individuals* or *social conditions* are more to blame for crime and lawlessness in the United States?" The response choices themselves should be phrased to make each one socially acceptable. It is a good idea to replace questions with Likert-style response choices that focus on the word *agree* (previous section) with others. For example, "To what extent do you support or oppose the new health care plan?" (options range from "strongly support" to "strongly oppose") is a better approach than the question "To what extent do you agree or disagree with the statement: 'The new health care plan is worthy of support'?" (response choices range from "strongly agree" to "strongly disagree"). For the same reason, simple true–false and yes–no response choices should be avoided (Saris et al. 2010; Schaeffer and Presser 2003:80–81).

(previous section)(previous section)You may also gain a more realistic assessment of respondents' sentiment by adding to a question a counterargument in favor of one side to balance an argument in favor of the other side. Thus, don't just ask in an employee survey whether employees should be required to join the union; instead, ask whether employees should be required to join the union or be able to make their own decision about joining. In one survey, 10% more respondents said they favored mandatory union membership when the counterargument was left out than when it was included. It is reassuring to know, however, that this approach does not change the distribution of answers to questions about which people have very strong beliefs (Schuman and Presser 1981:186).

When an illegal or socially disapproved behavior or attitude is the focus, some respondents will be reluctant to agree that they have ever done or thought such a thing. In this situation, the goal is to write a question and response choices that make agreement seem more acceptable (Tourangeau, Rips, and Rasinski 2000:288). For example, Dillman (2000:75) suggests we ask, "Have you ever taken anything from a store without paying for it?" rather than "Have you ever shoplifted something from a store?" Asking about a variety of behaviors or attitudes that range from socially acceptable to socially unacceptable will also soften the impact of agreeing with those that are socially unacceptable.

Minimize Fence-Sitting and Floating

Two related problems in writing survey questions also stem from people's desire to choose an acceptable answer. There is no uniformly correct solution to these problems; researchers have to weigh the alternatives in light of the concept to be measured and whatever they know about the respondents.

Fence-sitters, people who see themselves as being neutral, may skew the results if you force them to choose between opposites. In most cases, about 10% to 20% of such respondents—those who do not have strong feelings on an issue—will choose an explicit middle, neutral alternative (Schuman and Presser 1981:161–178). Having an explicit neutral response option is generally a good idea: It identifies fence-sitters and tends to increase measurement reliability (Schaeffer and Presser 2003:78).

Even more people can be termed **floaters**: respondents who choose a substantive answer when they really don't know or have no opinion. A third of the public will provide an opinion on a proposed law they know nothing about if they are asked for their opinion in a closed-ended survey question that does not include "Don't know" as an explicit response choice, with that tendency more pronounced among those with less education. However, 90% of these persons will select the "Don't know" response if they are explicitly given that option. On average, offering an

explicit response option increases the "Don't know" responses by about a fifth (Schuman and Presser 1981:113–160).

Despite the prevalence of floating, people often have an opinion but are reluctant to express it. Most political pollsters use **forced-choice questions** without a "Don't know" option to encourage respondents to express an opinion.

> It would not be very instructive for pollsters . . . to allow large numbers of voters to claim they are undecided all through the election season. We would miss the dynamics of change, we would be unable to tell how well candidates were doing in response to events, and publicly released polls would be out of synchronization with private, campaign polls. (Newport 1992:A28)

On balance, survey experts recommend that most questions not include explicit "Don't know" or "No opinion" options (Krosnick 1999:558; Schaeffer and Presser 2003:80).

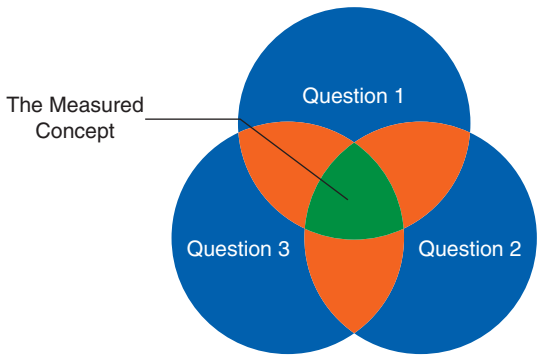
COMBINING QUESTIONS IN INDEXES

Writing single questions that yield usable answers is always a challenge. Simple though they may seem, single questions are prone to error because of **idiosyncratic variation**, which occurs when individuals' responses vary because of their reactions to particular words or ideas in the question. Differences in respondents' backgrounds, knowledge, and beliefs almost guarantee that some will understand the same question differently.

The guidelines in this chapter for writing clear questions should help reduce idiosyncratic variation caused by different interpretations of questions. But the best option is often to develop multiple questions about a concept and then to average the responses to those questions in a composite measure termed an *index* or *scale.* The idea is that idiosyncratic variation in particular responses will average out, so that the main influence on the combined measure will be the concept upon which all the questions focus. The index can be considered a more complete measure of the concept than can any one of the component questions.

Figure 8.5 illustrates an index in which the substantial area of overlap indicates the questions are measuring a common concept. Special statistics called *reliability measures* help researchers decide whether responses are consistent in this way. The most common of these, Cronbach's alpha (α), varies from 0 to 1. A score of 0 indicates that answers to different questions in the index are completely unrelated while a score of 1 indicates that the same response is given to every question in the index. An index is not considered sufficiently reliable unless α is at least.7. When

FIGURE 8.5 ■ Overlapping Dimensions of a Concept



an index score is used to make critical decisions, such as about employment or academic placement, an even higher level of reliability should be required (DeVellis 2017:146).

Because of the popularity of survey research, indexes already have been developed to measure many concepts, and some of these indexes have proved to be reliable in a range of studies. It usually is much better to use such an index to measure a concept than to try to devise questions to form a new index. Use of a preexisting index both simplifies the work involved in designing a study and facilitates comparison of findings to those obtained in other studies.

The questions in Figure 8.6 form an index used to measure the concept of depression and anxiety—the PHQ-4—that we used in our survey of Boston residents during the pandemic. Many researchers in different studies have found these questions form a reliable index. Note that these questions concern symptoms of depression and anxiety. People may have idiosyncratic reasons for having a particular symptom without being depressed or anxious; for example, a student waiting for an exam grade may be anxious, but just for that particular (idiosyncratic) reason. But by averaging the answers to questions about several symptoms, the index score reduces the impact of this idiosyncratic variation.

FIGURE 8.6 ■ Example of an Index: The Patient Health Questionnaire for Anxiety and Depression (PHQ-4)

43. During the past 2 weeks, how often have you been bothered by feeling down, depressed or hopeless?
    - ☐ Not at all
    - ☐ Several days
    - ☐ More than half the days
    - ☐ Nearly every day

44. During the past 2 weeks, how often have you been bothered by having little interest or pleasure in doing things?
    - ☐ Not at all
    - ☐ Several days
    - ☐ More than half the days
    - ☐ Nearly every day

45. During the past 2 weeks, how often have you been bothered by feeling nervous, anxious or on edge?
    - ☐ Not at all
    - ☐ Several days
    - ☐ More than half the days
    - ☐ Nearly every day

46. During the past 2 weeks, how often have you been bothered by not being able to stop or control worrying?
    - ☐ Not at all
    - ☐ Several days
    - ☐ More than half the days
    - ☐ Nearly every day

Source: Living in Boston During COVID-19 Survey. Data source: Center for Survey Research, University of Massachusetts Boston. June 2021. Questions from Kroenke, K., Spitzer, R. L., Williams, J. B., Löwe, B. 2009. "An ultra-brief screening scale for anxiety and depression: The PHQ-4." *Psychosomatics* 50(6):613–621.

Three cautions are in order:

1. *Our presupposition that each component question is indeed measuring the same concept may be mistaken.* Although we may include multiple questions in a survey to measure one concept, we may find that answers to the questions are not related to one another, and so the index cannot be created. Alternatively, we may find that answers to just a few of the questions are not related to the answers given to most of the others. We may therefore decide to discard these questions before computing the average that makes up the index.

2. *Combining responses to specific questions can obscure important differences in meaning among the questions.* My research on the impact of AIDS prevention education in shelters for the homeless provides an example. In this research, I asked a series of questions to ascertain respondents' knowledge about HIV risk factors and about methods of preventing exposure to those risk factors. I then combined these responses into an overall knowledge index. I was somewhat surprised to find the knowledge index scores were no higher in a shelter with an AIDS education program than in a shelter without such a program. However, further analysis showed that respondents in the shelter with an AIDS education program were more knowledgeable than the other respondents about the specific ways of preventing AIDS, which were, in fact, the primary focus of the program (Schutt, Gunston, and O'Brien1992).

3. *The questions in an index may cluster together in subsets.* All the questions may be measuring a common concept, but the concept may have several different aspects, or dimensions. A multidimensional index has been obtained in which answers to questions representing the different dimensions tend to be more similar than answers to questions representing other dimensions. This is represented by our activity index in the pandemic survey (Figure 8.3). The different behaviors capture the frequency of activities that broke social distancing rules during the pandemic (like attending a gathering with more than 10 people), some that were required to meet basic needs (like going to a food pantry), and some that reduced risk (such as food deliveries). By creating subscales for each of these dimensions, we distinguished types of behaviors that different social groups engaged in.

Another approach to creating an index score is to give different weights to the responses to different questions before summing or averaging the responses. Such a weighted index is also termed a *scale.* The scaling procedure might be as simple as arbitrarily counting responses to one question as worth two or three times as much as responses to another question, but most often, the weight applied to each question is determined through empirical testing. For example, based on Mooney and Lee's (1995) research on abortion law reform, the scoring procedure for a scale of support for abortion might give a 1 to agreement that abortion should be allowed "when the pregnancy resulted from rape or incest" and a 4 to agreement with the statement that abortion should be allowed "whenever a woman decided she wanted one." In other words, agreeing that abortion is allowable in any circumstances is much stronger support for abortion rights than is agreeing that abortion should be allowed in the case of rape or incest.

Another important decision in writing questions for an index is whether to reverse the wording of some questions so earlier responses refer to lower levels of the concept (while other questions are phrased so earlier responses refer to higher levels of the concept). The goal is to reduce agreement bias for the overall index. For example, in the next statement, from the National Longitudinal Survey of Youth (U.S. Bureau of Labor Statistics, 2022 Child Self-Administered supplement), the response of "Strongly Agree" supports treating girls and boys the same, while the same response to the following statement indicates support of unequal treatment of girls and boys:

How much do you agree or disagree with the following statements?

Girls and boys should be treated the same at school.

1 Strongly Agree
2 Agree

3 Disagree
4 Strongly Disagree

A girl should NOT let a boy know she is smarter than he is.

1 Strongly Agree
2 Agree
3 Disagree
4 Strongly Disagree

The numbers representing the responses to one subset are then reverse scored before the index is created, so a higher numerical value always means a higher value on the concept. Be aware, however, that the reliability of the overall index will be lower than if all the questions had been worded in the same direction in the first place.

DESIGNING QUESTIONNAIRES

Survey questions are answered as part of a **questionnaire** (or **interview schedule**, as it's often called in interview-based studies), not in isolation from other questions. The context created by the questionnaire has a major impact on how individual questions are interpreted and whether they are even answered. As a result, survey researchers must give very careful attention to the design of the questionnaire as well as to the individual questions that it includes.

There is no precise formula for identifying questionnaire features that reduce error. Nonetheless, some key principles should guide the design of any questionnaire, and some systematic procedures should be considered for refining it.

Build on Existing Instruments

Resources such as Miller and Salkind's (2002) *Handbook of Research Design and Social Measurement* can give you many ideas about existing instruments; your literature review at the start of a research project should be an even better source. The Roper Center's archives (www.ropercenter.cornell.edu) can be accessed from your university's library database (if it has subscribed) and contains over 650,000 questions and answers from surveys conducted since 1935 by over 150 survey firms.

But there is a trade-off here. Questions used previously may not concern quite the right concept or may not be appropriate in some ways to your population. A good rule of thumb is to consider a previously designed instrument if it measures the concept of concern to you and if you have no clear reason for thinking the instrument is not appropriate with your survey population.

Refine and Test Questions

Adhering to the preceding question-writing guidelines and considering questions used in previous surveys will go a long way toward producing a useful questionnaire. However, simply asking what appear to you to be clear questions does not ensure that people have a consistent understanding of what you are asking. You need some external feedback—the more of it, the better.

One important form of feedback results from simply discussing the questionnaire content with others. Persons who should be consulted include expert researchers, key figures in the locale or organization to be surveyed (e.g., elected representatives, company presidents, and community leaders), and individuals from the population to be sampled. Forming a panel of experts to

review the questions can also help: Presser and Blair (1994) recommend a panel of a psychologist, a questionnaire design expert, and a general methodologist (cited in Peterson 2000:116). Review of questions by an expert panel is the least expensive method and identifies the greatest number of problems with questions (Presser and Blair 1994).

Another form of feedback comes from guided discussions between potential respondents, called *focus groups*, to check for a consistent understanding of terms and to identify the range of events or experiences about which people will be asked to report. By listening to and observing the focus group discussions, researchers can validate their assumptions about what level of vocabulary is appropriate and what people are going to be reporting (Fowler 1995). (See Chapter 10 for more about this technique.)

Professional survey researchers also use a technique for improving questions called the **cognitive interview:** asking people to describe what they are thinking when they answer questions (Dillman 2000:66–67; Fowler 1995). The researcher asks a test question, then probes with follow-up questions about how the respondent understood one or more words in the question, how confusing it was, and so forth (Schaeffer and Presser 2003:82). If the individuals interviewed in this way reflect the population to be surveyed, the method identifies many problems with proposed questions.

In a traditional **survey pretest**, a questionnaire is administered to a small set of respondents (perhaps 15–25) who are like those who will be sampled in the planned survey. After the pretest, the researchers examine responses—and interviewer experiences for an in-person survey—and try to identify questions that caused problems. In the pretest version of a written questionnaire, you may include some space for individuals to comment on each key question or, with in-person interviews, audio-record the test interviews for later review.

The value of a pretest can be enhanced with **behavior coding**, in which a researcher observes the interviews or listens to recorded interviews and codes, according to strict rules, the number of times that difficulties occur with questions (Krosnick 1999: 541). Such difficulties include respondents asking for clarification and interviewers rephrasing questions rather than reading them verbatim (Presser and Blair 1994: 74–75). This information is then used to improve question wording and instructions for interviewers about reading the questions (Schaeffer and Presser 2003).

Add Interpretive Questions

A survey researcher can also include **interpretive questions** in the survey itself to help the researcher understand what the respondent meant by their response to particular questions. An example from a study of people with motor vehicle driving violations illustrates the importance of interpretive questions:

> When asked whether their emotional state affected their driving at all, respondents would reply that their emotions had very little effect on their habits. Then, when asked to describe the circumstances surrounding their last traffic violation, respondents typically replied, "I was mad at my girlfriend," or "I had a quarrel with my wife," or "We had a family quarrel," or "I was angry with my boss." (Labaw 1980:71)

Were these respondents lying in response to the first question? Probably not. More likely, they simply didn't interpret their own behavior in terms of general concepts such as *emotional state*. But their responses to the first question were likely to be misinterpreted without the further detail provided by answers to the second open-ended question.

Maintain Consistent Focus

Except in an omnibus survey, a survey should be guided by a clear conception of the research problem under investigation and the population to be sampled. Does the study seek to describe some phenomenon in detail, to explain some behavior, or to explore some type of social relationship? Until the research objective is formulated clearly, survey design cannot begin. Throughout the process of questionnaire design, this objective should be the primary basis for making decisions about what to include and exclude and what to emphasize or treat in a cursory fashion. Moreover, the questionnaire should be viewed as an integrated whole, in which each section and every question serves a clear purpose related to the study's objective and complements other sections or questions.

Of course, not even the best researcher can anticipate the relevance of every question. Many researchers try to avoid "missing something" by erring on the side of adding more questions (Labaw 1980:40), but this runs the risk of decreasing interest in the survey and so reducing the completion rate.

RESEARCH IN THE NEWS

In Phone Surveys, People Are Happy to Talk (and Keep Talking)

"A strange thing kept happening." When interviewers at the Siena College Research Institute conducted a phone survey of New Yorkers one month after the pandemic shutdown in March 2020, they found many people they reached wanted to keep talking, even after the interview questions ended. One woman began to weep after revealing she'd lost her job; another said, "the only good thing about her day had been venturing outside and seeing a single flower blooming." Courtney Kennedy, director of survey research for the Pew Research Center, also found that people "really appreciated the chance to talk about this." Response rates rose, reducing the gap in responding for younger people and for those without college degrees.

For Further Thought?

1. How do you think interviewers should respond when respondents express strong feelings in an interview? Explain your opinion.
2. What may the increased response rate in phone interviews during the early months of the pandemic indicate about the usual problem of low response rates? What does this experience suggest about how to improve response rates?

News source: Rossonello, Giovanni, and Sarah Lyall. 2021. "In Phone Surveys, People Are Happy to Talk (and Keep Talking)." *The New York Times*, April 18, A1, A21.

Order the Questions

The order in which questions are presented influences how respondents react to the questionnaire as a whole and how they may answer some questions (Schwarz 2010:47). As a first step, the individual questions should be sorted into broad thematic categories, which then become separate sections in the questionnaire. For example, our survey during the pandemic included sections about behavior during the pandemic, work situation, financial concerns, and feelings about COVID-19 guidelines (Center for Survey Research 2021). Both the sections and the

questions within the sections should be organized in a logical order that would make sense in a conversation.

The first question deserves special attention, particularly if the questionnaire is to be self-administered. This question signals to the respondent what the survey is about, whether it will be interesting, and how easy it will be to complete. For these reasons, the first question should connect to the primary purpose of the survey, be interesting and easy, and apply to everyone in the sample (Dillman 2000:92–94). Here is our first question in the Living in Boston During COVID-19 survey:

> How worried are you about you or someone in your family being infected with the COVID-19 coronavirus? . . .
>
> **Extremely worried**
> **Very worried**
> **Somewhat worried**
> **Not too worried**
> **Not at all worried**

Question order can lead to **context effects** when one or more questions influence how subsequent questions are interpreted (Schober 1999:88–89). For example, when a sample of the general public was asked, "Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children?" fifty-eight percent said yes. However, when this question was preceded by a less permissive question that asked whether the respondent would allow abortion of a defective fetus, only 40% said yes. Asking the question about a defective fetus altered respondents' frame of reference, perhaps by making abortion simply to avoid having more children seem frivolous by comparison (Turner and Martin 1984:135). Context effects have also been identified in the measurement of general happiness, in what is termed a **part–whole question effect** (Peterson 2000:113). Married people tend to report they are happier "in general" if the general happiness question is preceded by a question about their happiness with their marriage (Schuman and Presser 1981:23–77). The impact of question order tends to be greater for general, summary-type questions like this.

Some questions may be presented in a *matrix* format. **Matrix questions** are a series of questions that concern a common theme and that have the same response choices. The questions are written so a common initial phrase applies to each one, as in Figure 8.3. This format shortens the questionnaire by reducing the number of words that must be used for each question. It also emphasizes the common theme among the questions and so invites answering each question in relation to other questions in the matrix. It is a good idea to provide an explicit instruction to "Check one response on each line" in a matrix question because some respondents will think they have completed the entire matrix after they have responded to just a few of the specific questions.

Make the Questionnaire Attractive

An attractive questionnaire is more likely to be completed and less likely to confuse either the respondent or, in an interview, the interviewer. Printing a multipage questionnaire in booklet form usually results in the most attractive and simple-to-use questionnaire. Printing on both sides of folded-over legal-size paper (8½" by 14") is a good approach (Dillman 2000:80–86). An attractive questionnaire does not look cramped; plenty of white space—more between questions than within question components—makes the questionnaire appear easy to complete. Response choices are distinguished clearly and consistently, perhaps by formatting them with light print

(while questions are formatted with dark print) and keeping them in the middle of the pages. Response choices are listed vertically rather than horizontally across the page.

The proper path through the questionnaire for each respondent is identified with arrows or other graphics and judicious use of spacing and other aspects of layout. Respondents should not be confused about where to go next after they are told to skip a question. Instructions should help route respondents through skip patterns, and such skip patterns should be used infrequently. Instructions should also explain how each type of question is to be answered (e.g., by circling a number or writing a response) in a neutral way that isn't likely to influence responses. Some distinctive formatting should be used to identify instructions.

Consider Language and Culture

Should the survey be translated into one or more languages? In the 21st century, no survey plan can be considered complete until this issue has been considered. In the United States in 2019, 13.7% of all persons were foreign born, and 8.2% of those age 5 or older spoke English "not very well" (U.S. Census Bureau 2020, Selected Social Characteristics in the United States, Table DP02: https://data.census.gov/cedsci/table?q = foreign%20born&tid = ACSDP1Y2019.DP02).

When those who are not fluent in English are a sizable portion of a population, omitting them from a survey can result in a misleading description of the population. So survey researchers find increasingly that they must translate their questionnaires into one or more languages to represent the population of interest. This does not simply mean picking up a bilingual dictionary, clicking "translate" in a web browser, or hiring a translator to translate the questions and response choices word for word. The U.S. Census Bureau's (2006) guidelines for translation designate the literal translation as only one step in the process. What is needed is to achieve some equivalence of the concepts in different cultures (Church 2010:154–159). The U.S. Census Bureau and the World Health Organization (n.d.) recommend that questionnaires be translated by a team that includes trained translators, persons who are specialists in the subject matter of the survey, persons with expertise in questionnaire design, and experts with several of these skills who can review the translation and supervise a pretest (Pan and de la Puente 2005). Cultural competence requires knowledge of stereotypes to avoid insensitivity to cultural norms (Robinson and Leonard 2019).

ORGANIZING SURVEYS

There are five basic social science survey designs: (1) mailed, (2) group-administered, (3) phone, (4) in-person, and (5) electronic. Survey researchers can also combine elements of two or more of these basic designs in mixed-mode surveys. Table 8.1 summarizes the typical features of the five basic survey designs.

- *Manner of administration.* The five survey designs differ in the way the questionnaire is administered (see Table 8.1). Mailed, group, and **electronic surveys** are completed by the respondents themselves. During phone and in-person interviews, the researcher or a staff person asks the questions and records the respondent's answers. However, new mixed-mode surveys break down these distinctions. For example, in audio computer-assisted self-interviewing (or audio-CASI), the interviewer gives the respondent a laptop and a headset (Tourangeau 2004:790–791). The respondent reads the questions on the computer screen, hears the questions in the headset, and responds by choosing answers on the computer screen.

TABLE 8.1 ■ Typical Features of the Five Survey Designs

Design	Manner of Administration	Setting	Questionnaire Structure	Cost
Mailed survey	Self	Individual	Mostly structured	Low to moderate
Group survey	Self	Group	Mostly structured	Very low
Phone survey	Professional	Individual	Structured	Moderate
In-person interview	Professional	Individual	Structured or unstructured	High
Electronic survey	Self	Individual	Mostly structured	Low

- *Setting.* Most surveys are conducted in settings where only one respondent completes the survey at a time; most mail and electronic questionnaires and phone or in-person interviews are intended for completion by only one respondent at a time. A variant of the standard survey is a questionnaire distributed simultaneously to a group of respondents (such as students or employees), who complete the survey while the researcher (or assistant) waits.
- *Questionnaire structure.* Survey designs also differ in the extent to which the researcher structures the content and order of questions in advance. Most mailed, group, phone, and electronic surveys are highly structured, fixing in advance the content and order of questions and response choices. Some of these types of surveys, particularly mailed surveys, may include open-ended questions (respondents write in their answers rather than checking off one of several response choices). In-person interviews are often highly structured, but they may also include many questions without fixed response choices. Moreover, some interviews may proceed from an interview guide rather than a fixed set of questions. In these relatively unstructured interviews, the interviewer covers the same topics with respondents but varies questions according to the respondent's answers to previous questions. Extra questions are added as needed to clarify or explore answers to the most important questions. Computers make it easy for researchers to use complex branching patterns in questionnaires administered in person, on the phone, or on the web because the computer can present different questions based on responses to prior questions (Tourangeau 2004:789).
- *Cost.* In-person interviews are the most expensive type of survey. Phone interviews are much less expensive, but the low rates of response in recent years makes this approach inadvisable for surveys of the general population and many segments of it. Surveying by mail is cheaper yet. Electronic surveys can be the least expensive method because there are no interviewer costs, no mailing costs, and, for many designs, almost no costs for data entry. However, extra staff time and programming expertise are required to prepare an electronic questionnaire (Tourangeau, Conrad, and Couper 2012).

Mailed, Self-Administered Surveys

A **mailed survey** is conducted by mailing a questionnaire to respondents, who then administer the survey themselves. The central concern in a mailed survey is maximizing the response rate. Even an attractive questionnaire full of clear questions will probably be returned by no more

than 30% of a sample unless extra steps are taken to increase the rate of response. In the language of *social exchange theory*, the costs of responding are perceived to be much higher than any anticipated rewards for doing so. Fortunately, the conscientious use of a systematic survey design with multiple follow-ups will often lead to a response rate of 70% or higher rate (Dillman 2000).

Sending follow-up mailings to nonrespondents is the single most important requirement for obtaining an adequate response rate to a mailed survey. The follow-up mailings explicitly encourage initial nonrespondents to return a completed questionnaire; implicitly, they convey the importance of the effort. Dillman (2000:155–158, 177–188) has demonstrated the effectiveness of a standard procedure for the mailing process:

- A few days before the questionnaire is to be mailed, send a brief letter to respondents that notifies them of the importance of the survey they are to receive.
- Send the questionnaire with a well-designed, personalized cover letter (see the following description), a self-addressed, stamped return envelope, and, if possible, a token monetary reward. The materials should be inserted in the mail-out envelope so they will all be pulled out together when the envelope is opened (Dillman 2000: 174–175). There should be no chance that the respondent will miss something.
- Send a reminder postcard, thanking respondents and reminding nonrespondents, to all sample members two weeks after the initial mailing. The postcard should be friendly in tone and must include a phone number for those people who may not have received the questionnaire. It is important that this postcard be sent before most nonrespondents will have discarded their questionnaire, even though this means the postcard will arrive before all those who might have responded to the first mailing have done so.
- Send a replacement questionnaire with a new cover letter only to nonrespondents, two to four weeks after the initial questionnaire mailing. This cover letter should be a bit shorter and more insistent than the original cover letter. It should note that the recipient has not yet responded, and it should stress the survey's importance. Of course, a self-addressed, stamped return envelope must be included.
- The final step is taken six to eight weeks after the initial survey mailing. This step uses a different mode of delivery (either priority or special delivery) or a different survey design—usually an attempt to administer the questionnaire over the phone. These special procedures emphasize the importance of the survey and encourage people to respond.

The **cover letter** for a mailed questionnaire is critical to the success of a mailed survey. This statement to respondents sets the tone for the questionnaire. A carefully prepared cover letter should increase the response rate and result in more honest and complete answers to the survey questions; a poorly prepared cover letter can have the reverse effects.

The cover letter or introductory statement must be

- *Credible.* The letter should establish that the research is being conducted by a researcher or organization the respondent is likely to accept as a credible, unbiased authority. According to one investigation, a sponsor known to respondents may increase the rate of response by as much as 17%. Government sponsors tend to elicit high rates of response. Research conducted by well-known universities and recognized research organizations (e.g., Gallup or RAND) is also usually credible in this sense. The next most credible sponsors are state headquarters of an organization and then other people

in a similar field. Publishing firms, students (sorry!), and private associations elicit the lowest response rates.

- *Personalized.* The cover letter should include a personalized salutation (using the respondent's name, e.g., not just "Dear Student"), close with the researcher's signature (blue ballpoint pen is best because that makes it clear the researcher has personally signed), and refer to the respondent in the second person ("Your participation...").
- *Interesting.* The statement should interest the respondent in the contents of the questionnaire. Never make the mistake of assuming that what is of interest to you will also interest your respondents.
- *Responsible.* Reassure the respondent that the information you obtain will be treated confidentially and include a phone number to call if the respondent has any questions or would like a summary of the final report. Point out that the respondent's participation is completely voluntary (Dillman 1978:165–172).

Figure 8.7 is an example of a cover letter for a questionnaire.

FIGURE 8.7 ■ Sample Questionnaire Cover Letter

University of Massachusetts Boston
Department of Sociology

Jane Doe
AIDS Coordinator
Shattuck Shelter

Dear Ms. Doe:

AIDS is an increasing concern for homeless people and for homeless shelters. The enclosed survey is about the AIDS problem and related issues confronting shelters. It is sponsored by the Life Lines AIDS Prevention Project for the Homeless—a program of the Massachusetts Department of Public Health.

As an AIDS coordinator/shelter director, you have learned about homeless persons' problems and about implementing programs in response to those problems. The Life Lines Project needs to learn from your experience. Your answers to the questions in the enclosed survey will improve substantially the base of information for improving AIDS prevention programs.

Questions in the survey focus on AIDS prevention activities and on related aspects of shelter operations. It should take about 30 minutes to answer all the questions.

Every shelter AIDS coordinator (or shelter director) in Massachusetts is being asked to complete the survey. And every response is vital to the success of the survey: The survey report must represent the full range of experiences.

You may be assured of complete confi dentiality. No one outside of the university will have access to the questionnaire you return. (The ID number on the survey will permit us to check with nonrespondents to see if they need a replacement survey or other information.) All information presented in the report to Life Lines will be in aggregate form, with the exception of a list of the number, gender, and family status of each shelter's guests.

Please mail the survey back to us by Monday, June 4, and feel free to call if you have any questions.

Thank you for your assistance.

Yours sincerely,

Russell K. Schutt

Russell K. Schutt, PhD
Project Director

Stephanie Howard

Stephanie Howard
Project Assistant

Other steps are necessary to maximize the response rate, such as the following (Fowler 1988:99–106; Mangione 1995:79–82; Miller and Salkind 2002:164):

- Enclose a token incentive with the survey. A $2 or $5 bill seems to be the best incentive. It is both a reward for the respondent and an indication of your trust that the respondent will carry out their end of the "bargain." The response rate to mailed surveys increases by 19 percentage points, on average, in response to such an incentive (Church 1993). Offering electronic money, a redeemable voucher, or a donation to a charity is less effective than providing cash (Toepoel 2016:109–113). Offering a large monetary reward or some type of lottery ticket only for those who return their questionnaire is actually less effective, apparently because it does not indicate trust in the respondent (Dillman 2000:167–170).
- Include a stamped, self-addressed return envelope with each copy of the questionnaire. This reduces the cost for responding. The stamp helps personalize the exchange and is another indication of trust in the respondent (who could use the stamp for something else). Using a stamp rather than metered postage on the mail-out envelope does not seem to influence the response rate, but it is very important to use first-class rather than bulk-rate postage (Dillman 2000:171–174).

Related to the threat of nonresponse in mailed surveys is the hazard of incomplete responses. Some respondents may skip some questions or just stop answering questions at some point in the questionnaire. Some surveys now use special questions to identify respondents who are simply inattentive and so are more likely to give a "no opinion" type of response (Alvarez et al. 2019). Fortunately, potential respondents who have decided to participate in a survey usually complete it.

Group-Administered Surveys

A **group-administered survey** is completed by individual respondents assembled in a group. The response rate is not usually a major concern in surveys that are distributed and collected in a group setting because most group members will participate. The real difficulty with this method is that it is seldom feasible because it requires what might be called a captive audience, such as students, employees, members of the armed forces, and some institutionalized populations.

Whoever is responsible for administering the survey to the group must be careful to minimize comments that might bias answers or that could vary between different groups in the same survey (Dillman 2000:253–256). In addition to a written cover letter with the questionnaire, an introductory statement should be read to the group that expresses appreciation for their participation, describes the steps of the survey, and emphasizes that the survey is not the same as a test. To emphasize confidentiality, respondents should be given an envelope in which to seal their completed questionnaires.

Another issue of special concern with group-administered surveys is the possibility that respondents will feel coerced to participate and, as a result, will be less likely to answer questions honestly. No complete solution to this problem exists, but it helps to emphasize in the introductory statement the researcher's independence and to give participants a chance to ask questions about the survey. The sponsor should also understand the need to keep a low profile and to allow the researcher both control over the data and autonomy in report writing. Participation in group-administered surveys of grade school and high school students also requires parental permission, which can reduce participation.

Telephone Surveys

In a **phone survey**, interviewers question respondents over the phone and then record respondents' answers. Phone interviewing became a popular method of conducting surveys in the United States because almost all families had phones by the latter part of the 20th century.

Reaching Sample Units

Because published telephone directories leave out many households, most telephone surveys use random digit dialing at some point in the sampling process (Lavrakas 1987). A machine calls random phone numbers within the designated exchanges, whether or not the numbers are published. When the machine reaches an inappropriate household (such as a business in a survey that is directed to the general population), the phone number is simply replaced with another. Most survey research organizations use special methods to identify sets of phone numbers that are likely to include working numbers and so make the random digit dialing more efficient (Tourangeau 2004:778–780). However households are contacted, the interviewers must ask a series of questions at the start of the survey to ensure they are speaking to the appropriate member of the household (Ross 1990:7).

Maximizing Response to Phone Surveys

Four issues require special attention in phone surveys. First, because people often are not home, multiple callbacks will be needed for many sample members. Those with more money and education are more likely to be away from home; such persons are more likely to vote Republican, so the results of political polls can be seriously biased if few callback attempts are made (Kohut 1988).

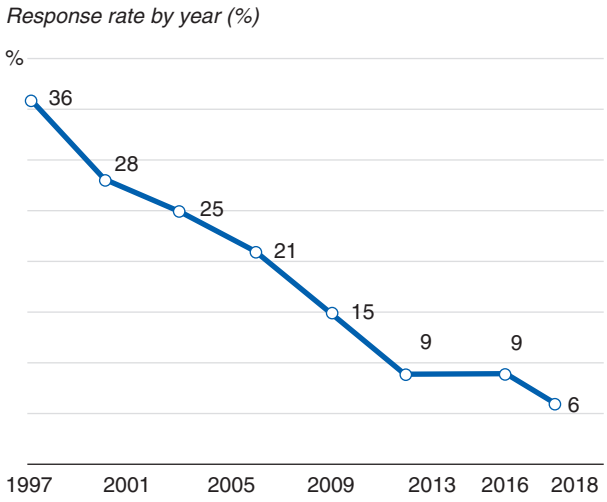
This problem has been compounded in recent years by social changes that are lowering the response rate in phone surveys (Tourangeau 2004:781–783) (see Figure 8.8). The Pew Research Center reports a decline in the response rate based on all those sampled from 36% in 1997 to only 6% in 2018 (Kennedy and Hartig 2019).

The number of callbacks needed to reach respondents by telephone has increased greatly in the past 20 years, with increasing numbers of single-person households, dual-earner families, and out-of-home activities. Survey research organizations have increased the usual number of phone contact attempts to 20, from between 4 and 8. The growth of telemarketing has created another problem for telephone survey researchers: Individuals have become more accustomed to "just say no" to calls from unknown individuals and organizations or to simply use their answering machines or caller ID to screen out unwanted calls (Dillman 2000:8, 28). Cell phone users are also harder (and more costly) to contact in phone surveys because their numbers are not in published directories. Households with a cell phone but no landline tend to be less educated, so telephone surveys overrepresent college graduates (Keeter et al. 2017) (see Figure 8.9).

The second issue researchers using phone surveys must cope with are difficulties because of the impersonal nature of phone contact. Visual aids cannot be used, so the interviewer must be able to convey verbally all information about response choices and skip patterns. Instructions to the interviewer must clarify how to ask each question, and response choices must be short.

Third, interviewers must be prepared for distractions because the respondent may be interrupted by other household members. Sprinkling interesting questions throughout the questionnaire may help maintain respondent interest. In general, rapport between the interviewer and the respondent is likely to be lower with phone surveys than with in-person interviews, and so respondents may tire and refuse to answer all the questions (Miller and Salkind 2002:317).

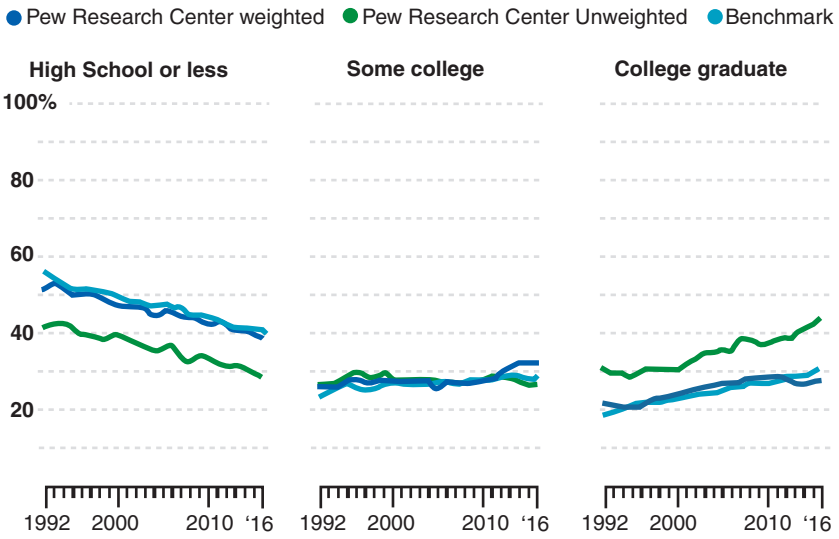
FIGURE 8.8 ■ Phone Survey Response Rates by Year, 1997–2018



Note: Response rate is AAPOR RR3. Only landlines sampled 1997–2006. Rates are typical for surveys conducted in each year.

Source: "Response Rates in Telephone Surveys Have Resumed Their Decline." Courtney Kennedy and Hannah Hartig. Pew Research Center, Washington, DC. (February 27, 2019). https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/.

FIGURE 8.9 ■ Education and Response to Phone Surveys



Source: "What Low Response Rates Mean for Telephone Surveys." Scott Keeter, Nick Hatley, Courtney Kennedy and Arnold Lau. Pew Research Center, Washington, DC. (May 15, 2017). https://www.pewresearch.org/methods/2017/05/15/what-low-response-rates-mean-for-telephone-surveys/.

Distractions are a special problem when respondents are called on a cell phone because they could be driving, in a restaurant or other crowded area, at work, or otherwise involved in activities that make responding difficult and that would not occur in a survey using a landline in the home (AAPOR 2014).

The fourth special consideration for phone surveys is that careful interviewer training is essential. Procedures can be standardized more effectively, quality control maintained, and processing speed maximized when phone interviewers use **computer-assisted telephone interviews (CATI)**.

Computerized **interactive voice response (IVR)** survey technology allows even greater control over interviewer–respondent interaction. In an IVR survey, respondents receive automated calls and answer questions by pressing numbers on their touch-tone phones or speaking numbers that are interpreted by computerized voice recognition software. These surveys can also record verbal responses to open-ended questions for later transcription. Although they present some difficulties when many answer choices must be used or skip patterns must be followed, IVR surveys have been used successfully with short questionnaires and when respondents are highly motivated to participate (Dillman 2000:402–411). When these conditions are *not* met, potential respondents may be put off by the impersonality of this computer-driven approach.

Phone surveying had for decades been the method of choice for relatively short surveys of the general population. However, phone surveying can no longer be considered the best method to use for general-purpose surveys. You have already learned of the dramatic decline in phone survey response rates. Those who do respond are more likely to be engaged in civic issues than those who do not respond, so estimates of related attitudes and behaviors in phone surveys can be quite biased (Kohut et al. 2012). Response rates can be improved somewhat by extra effort and with special populations. In a phone survey of low-income women in a public health program (Schutt and Fawcett 2005), the University of Massachusetts Center for Survey Research achieved a 55.1% response rate from all eligible sampled clients after a protocol that included as many as 30 contact attempts; the response rate rose to 72.9% when it was calculated as a percentage of clients who were located (Roman 2005:7).

In-Person Interviews

What is unique to the **in-person interview**, compared with the other survey designs, is the face-to-face social interaction between interviewer and respondent. If money is no object, in-person interviewing is often the best survey design.

In-person interviewing has several advantages: Response rates are higher than with any other survey design; questionnaires can be much longer than with mailed, phone, or web-based surveys; the questionnaire can be complex, with both open-ended and closed-ended questions and frequent branching patterns; the order in which questions are read and answered can be controlled by the interviewer; the physical and social circumstances of the interview can be monitored; and respondents' interpretations of questions can be probed and clarified.

But researchers must be alert to some special hazards resulting from the presence of an interviewer. Respondents should experience the interview process as a personalized interaction with an interviewer who is very interested in the respondent's experiences and opinions. At the same time, however, every respondent should have the same interview experience—be asked the same questions in the same way by the same type of person, who reacts similarly to the answers (de Leeuw 2008:318). Therein lies the researcher's challenge—to plan an interview process that will be personal and engaging and yet consistent and nonreactive (and to hire interviewers who can carry out this plan). Careful training and supervision are essential because small differences in intonation or emphasis on particular words can alter respondents' interpretations of questions' meaning (Groves 1989:404–406; Peterson 2000:24). Without a personalized approach, the rate of response will be lower and answers will be less thoughtful—and potentially less valid. Without a consistent approach, information obtained from different respondents will not be comparable—it will be less reliable and less valid.

As with phone interviewing, computers can be used to increase control of the in-person interview. In a **computer-assisted personal interview (CAPI)** project, interviewers carry a laptop computer that is programmed to display the interview questions and to process the responses the interviewer types in, as well as to check that these responses fall within allowed ranges (Tourangeau 2004:790–791). Interviewers seem to like CAPI, and the data obtained are comparable in quality to data obtained in a noncomputerized interview (Shepherd et al. 1996). A CAPI approach also makes it easier for the researcher to develop skip patterns and experiment with different types of questions for different respondents without increasing the risk of interviewer mistakes (Couper et al. 1998).

The presence of an interviewer may make it more difficult for respondents to give honest answers to questions about socially undesirable behaviors, such as drug use, sexual activity, and not voting (Schaeffer and Presser 2003:75). CAPI is valued for this reason because respondents can enter their answers directly into the laptop without the interviewer knowing what their response is. Alternatively, interviewers can simply hand respondents a separate self-administered questionnaire containing the more sensitive questions. After answering these questions, the respondent seals the separate questionnaire in an envelope so the interviewer does not know the answers. When this approach was used for the GSS questions about sexual activity, about 21% of men and 13% of women who were married or had been married admitted to having cheated on a spouse ("Survey on Adultery" 1993:A20). The degree of rapport becomes a special challenge when survey questions concern issues related to such demographic characteristics as race or gender (Groves 1989). If the interviewer and respondent are similar on the characteristics at issue, the responses to these questions may differ from those that would be given if the interviewer and respondent differ on these characteristics. For example, a white respondent may not disclose feelings of racial prejudice to a Black interviewer that he would admit to a white interviewer.

Although in-person interview procedures are typically designed with the expectation that the interview will involve only the interviewer and the respondent, one or more other household members are often within earshot. In a mental health survey in Los Angeles, for example, almost half the interviews were conducted in the presence of another person (Pollner and Adams 1994). It is reasonable to worry that this third-party presence will influence responses about sensitive subjects. However, there is no consistent evidence that respondents change their answers because of the presence of another person.

Maximizing Response to Interviews

Several factors affect the response rate in interview studies. Contact rates tend to be lower in central cities, partly because of difficulties in finding people at home and gaining access to high-rise apartments and partly because of interviewer reluctance to visit some areas at night when people are more likely to be home (Fowler 1988:45–60). Single-person households also are more difficult to reach, whereas households with young children or elderly adults tend to be easier to contact (Groves and Couper 1998:119–154).

Refusal rates vary with some respondents' characteristics. People with less education participate somewhat less in surveys of political issues (perhaps because they are less aware of current political issues) (Groves 1989). High-income persons tend to participate less in surveys about income and economic behavior (perhaps because they are suspicious about why others want to know about their situation). Unusual strains and disillusionment in a society, like those experienced in the last decade in the United States and other countries, can also undermine the general credibility of research efforts and the ability of interviewers to achieve an acceptable response rate. These problems can be lessened with an advance letter introducing the survey

project and by multiple contact attempts throughout the day and evening, but they cannot be avoided entirely (Fowler 1988:52–53; Groves and Couper 1998). Encouraging interviewers to make small talk to increase rapport and delay asking a potential respondent to participate may reduce the likelihood of a refusal (Maynard, Freese, and Schaeffer 2010:810).

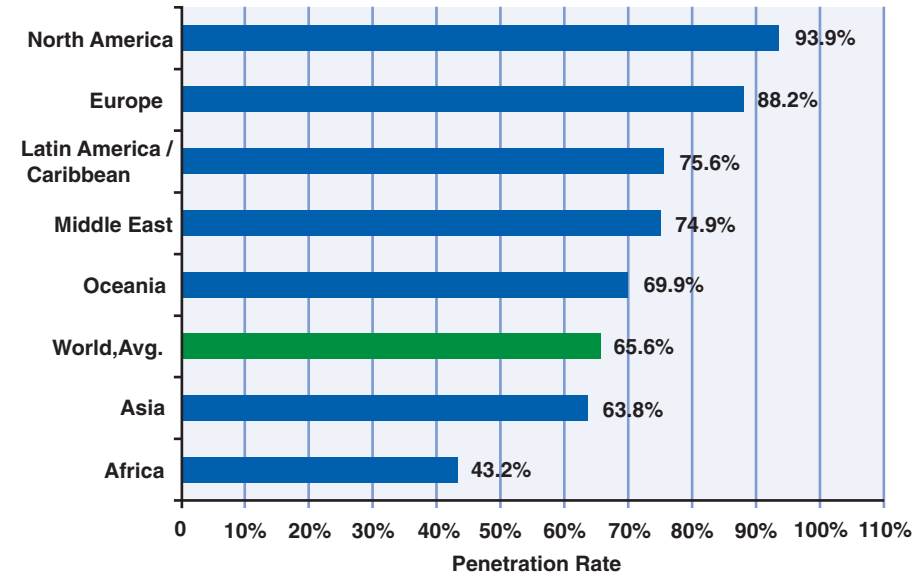
Web Surveys

Web surveys have become an increasingly useful survey method for two reasons: growth in the fraction of the population using the Internet and technological advances that make web survey design relatively easy and often superior to printed layouts. Many specific populations have very high rates of Internet use, so a web survey can be a good option for groups such as professionals, middle-class communities, members of organizations, and, of course, college students. Because of the Internet's global reach, web surveys also make it possible to survey multinational populations. However, coverage remains a major problem in many countries. All but 7% of North American households and 12% of European households were connected to the Internet in 2021, but rates of Internet usage are much lower in other parts of the world, ranging from about 75% in Latin America and the Middle East to 43.2% in Africa (see Figure 8.10; Internet World Statistics, 2021).

The extent to which the population of interest is connected to the web is the most important consideration when deciding whether to conduct a survey through the web. Other considerations that may increase the attractiveness of a web survey include the need for a large sample, for rapid turnaround, or for collecting sensitive information that respondents might be reluctant to discuss in person (Hewson et al. 2016:41–57; Sue and Ritter 2012:10–11).

There are several different approaches to engaging people in web surveys, each with unique advantages and disadvantages and somewhat different effects on the coverage problem. Many

FIGURE 8.10 ■ Worldwide Internet Penetration Rates by Region, 2021



Source: Internet World Stats. www.internetworkdstats.com.

web surveys begin with an e-mail message to potential respondents that contains a direct "hotlink" to the survey website (Gaiser and Schreiner 2009:70). It is important that such e-mail invitations include a catchy phrase in the subject line as well as attractive and clear text in the message itself (Sue and Ritter 2012:110–114). This approach is particularly useful when a defined population with known e-mail addresses is to be surveyed. The researcher can then send e-mail invitations to a representative sample without difficulty. To ensure the appropriate people respond to a web survey, researchers may require that respondents enter a personal identification number (PIN) to gain access to the web survey (Dillman 2000:378; Sue and Ritter 2012:103–104).

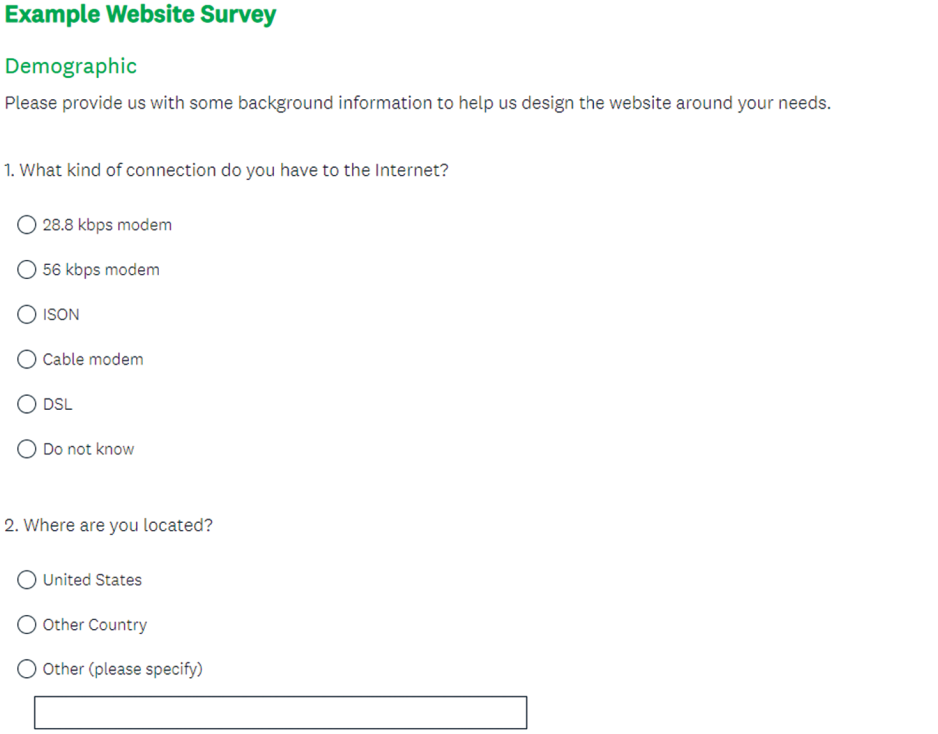
However, lists of unique e-mail addresses for the members of defined populations generally do not exist outside of organizational settings. Many people have more than one e-mail address, and often there is no apparent link between an e-mail address and the name or location of the person to whom it is assigned. It is not possible to ensure that respondents correctly identify their own characteristics (Toepoel 2016:49). As a result, there is no available method for drawing a random sample of e-mail addresses for people from any general population, even if the focus is only on those with Internet access (Dillman 2007:449). Instead, web survey participants should be recruited from mailings or phone calls to their home addresses, with the web survey link sent after they have agreed to participate (Toepoel 2016:72–73). If individuals or families that are willing to participate but do not have Internet access are provided with a computer (if needed) and an Internet connection, sampling bias can be reduced—and the extra costs can be spread out over time when these respondents agree to participate in future surveys. Gfk Knowledge Networks is a company that received funding from the U.S. National Science Foundation to create such a web survey panel (Couper and Miller 2008:832–833). However, many persons will still not agree to participate. and panel attrition can create coverage bias for new surveys (Couper 2000:485–489; Frankel and Hillygus 2013).

When web surveys are administered to "volunteer samples" that are recruited on a particular website or through social media, the result can be a very large but very biased sample of the larger population (Couper 2000:486–487; Dillman 2000:355). For example, the volunteer sample that participated in the National Geographic Society's global Survey 2000 was very large (50,000 respondents), but it represented the type of people who visit the society's website (middle class, young, North American) (Witte Amoroso and Howard 2000).

Bias in volunteer samples can be reduced somewhat by requiring participants to meet certain inclusion criteria (Selm and Jankowski 2006:440). Another approach to reducing coverage bias with volunteer samples recruited on the Internet is to weight respondents based on key characteristics so the resulting sample is more comparable to the general population in terms of such demographics as gender, race, age, and education (Couper and Miller 2008:832–833). It appears that weighting can reduce coverage bias by 30% to 60% (Tourangeau et al. 2012).

In contrast to problems of coverage, web surveys have some unique advantages for increasing measurement validity (Selm & Jankowski, 2006; Tourangeau et al., 2012). Questionnaires completed on the web can elicit more honest reports about socially undesirable behavior or experiences, including illicit behavior and victimization in the general population and failing course grades among college students, when compared with results with phone interviews (Kreuter, Presser, and Tourangeau 2008; Onoye, Goebert and Nishimura 2012; Parks, Pardi, and Bradizza 2006). Although web surveys should be short—with about 25–30 questions requiring no more than 10 minutes to complete—they are relatively easy to complete because respondents simply click on response boxes (see Figure 8.11). The survey can be programmed to move respondents easily through sets of questions (Hewson et al. 2016:142; Kreuter et al. 2008; Toepoel 2016:35).

FIGURE 8.11 ■ Survey Monkey Web Survey Example


Example Website Survey
Demographic
Please provide us with some background information to help us design the website around your needs.
1. What kind of connection do you have to the Internet?
28.8 kbps modem
56 kbps modem
ISON
Cable modem
DSL
Do not know
2. Where are you located?
United States
Other Country
Other (please specify)

Use of the visual, interactive web medium can also help. Pictures, sounds, and animation can be used as a focus of particular questions, and graphic and typographic variation can be used to enhance visual survey appeal. Definitions of terms can also "pop up" when respondents scroll over them (Dillman 2007:458–459). Responses can quickly be checked to make sure they fall within the allowable range. Because answers are recorded directly in the researcher's database, data entry errors are almost eliminated, and results can be reported quickly. SurveyMonkey (www.surveymonkey.com) and Qualtrics (www.qualtrics.com) are popular options for designing questionnaires and administering surveys online.

At this point, there is reason enough to consider the option of a web survey for many investigations, but also clear evidence of the need to consider carefully strengths and weaknesses of this approach before deciding to conduct a web survey. When people are sent a mailed survey that also provides a link to a web survey alternative, they overwhelmingly choose the paper survey (Couper 2000:488),as occurred in our Boston pandemic survey. Surveys by phone continue to elicit higher rates of response than web surveys of a defined population (Kreuter et al. 2008).

Mixed-Mode Surveys

Mixed-mode surveys allow the strengths of one survey design to compensate for the weaknesses of another and can maximize the likelihood of securing data from different types of respondents (Dillman 2007:451–453; Selm and Jankowski 2006). For example, a survey may be sent electronically to sample members who have e-mail addresses and mailed to those who don't. Phone reminders may be used to encourage responses to web or paper surveys, or a letter of introduction

may be sent in advance of calls in a phone survey (Guterbock 2008). In one comparative study, the response rate to a telephone survey rose from 43% to 80% when it was followed by a mailed questionnaire (Dillman 2007:456). Alternatively, nonrespondents in a mailed survey may be interviewed in person or over the phone.

The mixed-mode approach is not a perfect solution. Respondents to the same question may give different answers because of the survey mode rather than because they actually have different opinions (Rookey, Hanway, and Dillman 2008:974; Toepoel 2016:98–99). Use of the same question structures, response choices, and skip instructions across modes substantially reduces the likelihood of mode effects, as does using only a small number of response choices for each question (Dillman 2000:232–240; Dillman and Christian 2005).

A Comparison of Survey Designs

Which survey design should be used when? Group-administered surveys are similar, in most respects, to mailed surveys, except they require the unusual circumstance of having access to the sample in a group setting. We therefore don't need to consider this survey design by itself; what applies to mailed surveys applies to group-administered survey designs, with the exception of sampling issues. The features of mixed-mode surveys depend on the survey types that are being combined. Thus, we can focus our comparison on the four survey designs that involve the use of a questionnaire with individuals sampled from a larger population: (1) mailed surveys, (2) phone surveys, (3) in-person surveys, and (4) electronic surveys. Table 8.2 summarizes their strong and weak points.

TABLE 8.2 ■ Advantages and Disadvantages of the Four Survey Designs

Characteristics of Design	Mail Survey	Phone Survey	In-Person Survey	Web Survey
Representative sample				
Opportunity for inclusion is known				
For completely listed populations	High	High	High	Medium
For incompletely listed populations	Medium	Medium	High	Low
Selection within sampling units is controlled (must respond, e.g., specific family members)	Medium	High	High	Low
Respondents are likely to be located				
If samples are heterogeneous	Medium	Medium	High	Low
If samples are homogeneous and specialized	High	High	High	High

(*Continued*)

TABLE 8.2 ■ Advantages and Disadvantages of the Four Survey Designs (*Continued*)

Characteristics of Design	Mail Survey	Phone Survey	In-Person Survey	Web Survey
Questionnaire construction and question design				
Allowable length of questionnaire	Medium	Medium	High	Low
Ability to include				
Complex questions	Medium	Low	High	High
Open questions	Low	High	High	Medium
Screening questions	Low	High	High	High
Tedious, boring questions	Low	High	High	Low
Ability to control question sequence	Low	High	High	High
Ability to ensure questionnaire completion	Medium	High	High	Low
Distortion of answers				
Odds of avoiding social desirability bias	High	Medium	Low	High
Odds of avoiding interviewer distortion	High	Medium	Low	High
Odds of avoiding contamination by others	Medium	High	Medium	Medium
Administrative goals				
Odds of meeting personnel requirements	High	High	Low	Medium
Odds of implementing quickly	Low	High	Low	High
Odds of keeping costs low	High	Medium	Low	High

Source: Adapted from Dillman 1978:74–75. *Mail and Telephone Surveys: The Total Design Method.* Reprinted by permission of John Wiley & Sons, Inc.

The most important consideration in comparing the advantages and disadvantages of the four methods is the likely response rate they will generate. In-person surveys are preferable in the possible length and complexity of the questionnaire itself, as well as with respect to the researcher's ability to monitor conditions while the questionnaire is completed. Mailed surveys often are preferable for asking sensitive questions, although this problem can be lessened in an interview by giving respondents a separate sheet to fill out or a laptop on which to enter their answers. Although interviewers may themselves distort results, either by changing the wording of questions or by failing to record answers properly, survey research organizations can reduce

this risk through careful interviewer training and monitoring. Some survey supervisors will have interviews recorded so they can review the dialogue between interviewers and respondents and provide feedback to the interviewers to help improve their performance. Some survey organizations have also switched to having in-person interviews completed entirely by the respondents on a laptop as they listen to prerecorded questions.

People interviewed by phone tend to be less interested in the survey than are those interviewed in person, so they tend to satisfice more—apparently in a desire to complete the survey more quickly— and they tend to be less trusting of the survey motives (Holbrook, Green, and Krosnick 2003).

The advantages and disadvantages of electronic surveys must be weighed considering the population that is to be surveyed and the capabilities at the time the survey is to be conducted. Currently, too many people lack Internet connections for survey researchers to use the Internet to survey the general population across the globe or to avoid a bias toward more educated, younger people even in the United States.

These various points about the different survey designs lead to two general conclusions. First, in-person interviews are the strongest design and generally preferable when sufficient resources and a trained interview staff are available; severely reduced response rates in telephone surveys make them no longer a viable alternative for many populations. Second, the "best" survey design for any study will be determined by the study's unique features and goals rather than by any absolute standard of what the best survey design is.

ETHICAL ISSUES IN SURVEY RESEARCH

Survey research usually poses fewer ethical dilemmas than do experimental or field research designs. Potential respondents to a survey can easily decline to participate, and a cover letter or introductory statement that identifies the sponsors of and motivations for the survey gives them the information required to make this decision. The methods of data collection are quite obvious in a survey, so little is concealed from the respondents. Only in group-administered surveys might the respondents be, in effect, a captive audience (probably of students or employees), and so these designs require special attention to ensure participation is truly voluntary. (Those who do not wish to participate may be told they can just hand in a blank form.)

The new proposed federal regulations to protect human subjects allow most survey research to be exempted from formal review (*Federal Register,* 2017:7261, 7262, 7264). Surveys fall within the exemption criteria that you learned about in Chapter 3, which stipulate that research is exempt from review if respondents cannot readily be identified (or procedures are in place to protect their privacy and maintain confidentiality of the data) or if disclosure of their responses would not place them at risk in terms of legal action, financial standing, employability, educational advancement, or reputation.

Confidentiality is most often the primary focus of ethical concern in survey research. Many surveys include essential questions that might, in some way, prove damaging to the subjects if their answers were disclosed. To prevent any possibility of harm to subjects because of the disclosure of such information, the researcher must preserve subject confidentiality. Nobody but research personnel should have access to information that could be used to link respondents to their responses, and even that access should be limited to what is necessary for specific research purposes. Respondents should be identified on the questionnaires only with numbers, and the names that correspond to these numbers should be kept in a safe location—unavailable to staff and others who might otherwise come across them. Follow-up mailings or contact

attempts that require linking the ID numbers with names and addresses should be carried out by trustworthy assistants under close supervision. For electronic surveys, encryption technology should be used to make information provided over the Internet secure from unauthorized persons.

Not many surveys can provide true **anonymity**, so that no identifying information is ever recorded to link respondents with their responses. The main problem with anonymous surveys is they preclude follow-up attempts to encourage participation by initial nonrespondents, and they prevent panel designs, which measure change through repeated surveys of the same individuals. In-person surveys rarely can be anonymous because an interviewer must, in almost all cases, know the name and address of the interviewee. To provide anonymity in a mail survey, the researcher should omit identifying codes from the questionnaire but could include a separate self-addressed, stamped postcard so the respondent can notify the researcher that the questionnaire has been returned without creating any linkage to the questionnaire itself (Mangione 1995:69).

Web surveys create some unique ethical issues. It is easy for respondents to skip through a consent form when responding to a survey online and so never to consider the issues. A good practice is to provide a list of statements about the project and require the respondent to check a box to indicate acceptance of each one (Hewson et al. 2016:103). Screening for eligibility is another problem heightened in web surveys. A survey intended for adults could as easily be completed by children; in fact, more than 80% of children lie about their age when using social media sites. Recruiting respondents at websites that cater only to adults will reduce the problem, as will clear guidelines about eligibility criteria. Clear instructions can also minimize the problem of multiple submissions of a survey via the Internet (Toepoel 2016:48–49).

CONCLUSIONS

Survey research is an exceptionally efficient and productive method for investigating a wide array of social research questions and doing so quickly, as was needed to inform responses to the pandemic. In addition to the potential benefits for social science, considerations of time and expense frequently make a survey the preferred data collection method. One or more of the six survey designs reviewed in this chapter (including mixed mode) can be applied to almost any research question. It is no wonder that surveys have become the most popular research method in sociology and that they frequently inform discussion and planning about important social and political questions.

The relative ease of conducting at least some types of survey research leads many people to imagine that special training and systematic procedures are not required. Nothing could be further from the truth. As a result of this widespread misconception, you will encounter a great many nearly worthless survey results. You must be prepared to examine carefully the procedures used in any survey before accepting its findings as credible. And if you decide to conduct a survey, you must be prepared to invest the time and effort that proper procedures require.

KEY TERMS

Anonymity (p. 294)
Behavior coding (p. 276)
Bipolar response options (p. 269)
Cognitive interview (p. 276)
Computer-assisted personal interview (CAPI) (p. 287)

Computer-assisted telephone interview (CATI)
Confidentiality (p. 293)
Context effects (p. 278)
Contingent question (p. 267)
Cover letter (p. 281)
Double negative
Double-barreled question
Electronic survey
Fence-sitters (p. 271)
Filter question
Floaters (p. 271)
Forced-choice questions (p. 272)
Group-administered survey (p. 283)
Idiosyncratic variation (p. 272)
In-person interview (p. 286)
Interactive voice response (IVR) (p. 286)
Interpretive questions (p. 276)
Interview schedule (p. 275)
Labeled unipolar response options (p. 269)
Likert item (p. 269)
Mailed survey (p. 280)
Matrix questions (p. 278)
Mixed-mode survey
Omnibus survey (p. 263)
Part–whole question effects
Phone survey (p. 284)
Questionnaire (p. 275)
Skip pattern
Social desirability bias (p. 269)
Split-ballot design (p. 263)
Survey pretest (p. 276)
Survey research (p. 261)
Unlabeled unipolar response options (p. 269)
Web survey

HIGHLIGHTS

LO 8.1 Summarize the reasons for the popularity of survey research.

- *Versatility:* Survey methods are versatile. A well-designed survey can enhance understanding of just about any social issue.
- *Efficiency:* Surveys also are popular because data can be collected from many people at relatively low cost and, depending on the survey design, relatively quickly.
- *Generalizability:* Survey methods lend themselves to probability sampling from large populations. Thus, survey research is appealing when sample generalizability is a central research goal. Surveys are also useful when cross-population generalizability is a key concern because they allow a range of social contexts and subgroups to be sampled.
- An omnibus survey shows just how versatile, efficient, and generalizable a survey can be. An omnibus survey covers a range of topics of interest to different social scientists, in contrast to the typical survey that is directed at a specific research question.

LO 8.2 Report the four types of errors of observation that must be minimized in a successful survey.

For a survey to succeed, it must minimize four types of error:

1. Poor measurement
2. Nonresponse
3. Inadequate coverage of the population
4. Sampling error

LO 8.3 Describe the basic principles for developing clear and meaningful survey questions.

- *Avoid confusing phrasing.* In most cases, a simple direct approach to asking a question minimizes confusion. One should present double-barreled questions nor include double negatives in survey questions.

- *Minimize the risk of bias*. Survey questions may trigger biases in respondents when they include biased or loaded words and phrases, response alternatives do not reflect the full range of possible sentiment on an issue, or if response alternatives on a continuum seem more attractive on one end of the continuum than the other.
- *Maximize the utility of response categories.* Response choices should be matched to the question they follow and reflect meaningful distinctions.
  - Questions with fixed response choices must provide one and only one possible response for everyone who is asked the question—that is, the response choices must be exhaustive and mutually exclusive.
  - Questions about thoughts and feelings will be more reliable if they refer to specific times or events.
  - Sometimes, problems with response choices can be corrected by adding questions. Adding questions may also improve memory about specific past events.
  - One common approach for measures of attitude intensity is to present a statement and then ask respondents to indicate their degree of agreement or disagreement.
  - "Check all that apply" and filter questions can be used as appropriate.
- *Avoid making either disagreement or agreement disagreeable to avoid social desirability bias.*
- *Avoid making either disagreement or agreement disagreeable to avoid social desirability bias.*
  - Present both sides of an attitude scale in a question and phrase response choices to make each one socially acceptable.
  - Consider replacing a range of response alternatives that focus on the word agree with other, more precise words.
  - You may gain a more realistic assessment of respondents' sentiment by adding to a question a counterargument in favor of one side to balance an argument in favor of the other side.
  - When an illegal or socially disapproved behavior or attitude is the focus, some respondents will be reluctant to agree that they have ever done or thought such a thing. In this situation, the goal is to write a question and response choices that make agreement seem more acceptable.
- *Minimize fence-sitting and floating.* Having an explicit neutral response option is generally a good idea to identify fence-sitters and increase measurement reliability. Meanwhile, it is generally a good practice to not include explicit "Don't know" or "No opinion" survey response options.

LO 8.4 Defend the rationale for using multi-question indexes.

- Multi-question indexes are advisable because single questions are prone to error due to idiosyncratic variation, which occurs when individuals' responses vary because of their reactions to particular words or ideas in the question as a result of their backgrounds, knowledge, and beliefs.
- To avoid idiosyncratic variation, the best option is often to develop multiple questions about a concept and then to average the responses to those questions in a composite measure termed an *index* or *scale.*

- The idea is that idiosyncratic variation in particular responses will average out, so that the main influence on the combined measure will be the concept upon which all the questions focus.
- Because of the popularity of survey research, indexes already have been developed to measure many concepts, and some of these indexes have proved to be reliable in a range of studies.
- Although multi-question indexes can be valuable, there are three cautions to keep in mind:
  - Assuming that each component question is indeed measuring the same concept may not be warranted.
  - Combining responses to specific questions can obscure important differences in meaning among the questions.
  - The questions in an index may cluster together in subsets. All the questions may be measuring a common concept, but the concept may have several different aspects, or dimensions.

LO 8.5 Appraise the best practices to reduce error through careful survey design.

- *Build on existing instruments*. A researcher should consider a previously designed instrument if it measures the concept of concern to them and if they have no clear reason for thinking the instrument is not appropriate with their survey population.
- *Refine and test questions*. A researcher can refine and test questions by discussing the questionnaire content with expert researchers, key figures at their research site, individuals from the sampled population, and focus groups of potential participants; by performing cognitive interviews; and by implementing a survey pretest with behavior coding.
- *Add interpretive questions.* Interpretive questions help the researcher understand what the respondent meant by their response to particular questions.
- *Maintain consistent focus.* Except in an omnibus survey, a survey should be guided by a clear conception of the research problem under investigation and the population to be sampled. This conception should be the primary basis for making decisions about what to include and exclude and what to emphasize or treat in a cursory fashion.
- *Order the questions.* The order in which questions are presented influences how respondents react to the questionnaire as a whole and how they may answer some questions.
- *Make the questionnaire attractive.* An attractive questionnaire is more likely to be completed and less likely to confuse either the respondent or, in an interview, the interviewer.
- *Consider language and culture.* Survey researchers find increasingly that they must translate their questionnaires into one or more languages to represent the population of interest.

LO 8.6 Compare the strengths and weaknesses of the five basic social science survey designs.

- There are five basic social science survey designs: (1) mailed, (2) group-administered, (3) phone, (4) in-person, and (5) electronic. Survey

researchers can also combine elements of two or more of these basic designs in mixed-mode surveys.

- Mailed surveys. The central concern in a mailed survey is maximizing the response rate. Even an attractive questionnaire full of clear questions will probably be returned by no more than 30% of a sample after one mailing. Sending follow-up mailings to nonrespondents is the single most important requirement for obtaining an adequate response rate to a mailed survey.
- Group-administered surveys. The response rate is not usually a major concern in surveys that are distributed and collected in a group setting because most group members will participate.
- Phone surveys. Severely reduced response rates in phone surveys make them no longer a viable alternative for many populations. To get an adequate response rate, multiple callbacks will be needed for many sample members.
- In-person surveys. Response rates are higher than with any other survey design; however, contact rates tend to be lower in central cities and among single-person households and refusal rates vary with some respondents' characteristics. Socially desirable responding is more of a problem than in other designs. In-person surveys are much more expensive than other survey designs. Careful interviewer training and supervision are required to ensure consistency.
- Electronic (web) surveys. The extent to which the population of interest is connected to the web affects web survey responses. When web surveys are administered to "volunteer samples" that are recruited on a particular website or through social media, the result can be a very large but very biased sample of the larger population. Web surveys allow great control over the order of questions and the use of visually attractive formats.

LO 8.7 Discuss the key ethical issues in survey research.

- Survey research usually poses fewer ethical dilemmas than do experimental or field research designs. Relatedly, the new proposed federal regulations to protect human subjects allow most survey research to be exempted from formal review.
- There are still two key ethical dilemmas to consider when conducting survey research:
  - Confidentiality: Many surveys include essential questions that might, in some way, prove damaging to the subjects if their answers were disclosed.
  - Anonymity: Not many surveys can provide true anonymity so that no identifying information is ever recorded to link respondents with their responses.
- Web surveys pose unique ethical issues. A web respondent may skip through a consent form and never consider the issues. Screening for eligibility is another problem heightened in web surveys.

DISCUSSION QUESTIONS

1. Response rates to phone surveys are declining, even as phone usage increases. Part of the problem is that lists of cell phone numbers are not available and wireless service providers may not allow outside access to their networks. Cell phone users may also have to pay for

incoming calls. Do you think regulations should be passed to increase the ability of survey researchers to include cell phones in their random digit dialing surveys? How would you feel about receiving survey calls on your cell phone? What problems might result from "improving" phone survey capabilities in this way?

2. In-person interviews have for many years been the "gold standard" in survey research because the presence of an interviewer increases the response rate, allows better rapport with the interviewee, facilitates clarification of questions and instructions, and provides feedback about the interviewee's situation. However, researchers who design in-person interviewing projects are now making increasing use of technology to ensure consistent questioning of respondents and to provide greater privacy for respondents answering questions. But having a respondent answer questions on a laptop while the interviewer waits is a very different social process than asking the questions verbally. Which approach would you favor in survey research? What trade-offs might there be in quality of information collected, rapport building, and interviewee satisfaction?

3. Each of the following questions was used in a survey that I received in the past. Evaluate each question and its response choices using the guidelines for writing survey questions presented in this chapter. What errors do you find? Try to rewrite each question to avoid such errors and improve question wording.
   - **a.** From an *InfoWorld* (computer publication) product evaluation survey:
     How interested are you in PostScript Level 2 printers?
     ___Very ____Somewhat ____Not at all
   - **b.** From the Greenpeace National Marine Mammal Survey: Do you support Greenpeace's nonviolent, direct action to intercept whaling ships, tuna fleets, and other commercial fishermen to stop their wanton destruction of thousands of magnificent marine mammals?
     ____Yes ____No ____Undecided
   - **c.** From a U.S. Department of Education survey of college faculty:
     How satisfied or dissatisfied are you with each of the following aspects of your instructional duties at this institution?

	Very Dissat.	Somewhat Dissat.	Somewhat Satisf.	Very Satisf.
a. The authority I have to make decisions about what courses I teach	1	2	3	4
b Time available for working with students as advisor, mentor	1	2	3	4

SPSS EXERCISES

What can we learn from the General Social Survey (GSS2020) data about the orientations of people who support capital punishment? Is it related to religion? Reflective of attitudes toward race? What about political views? Do attitudes and behavior concerning guns have some relation to support for capital punishment?

1. To answer these questions, we will use some version of each of the following variables in our analysis: PARTYID3, GOVAIDBLK3, FUND2020, OWNGUN_2, and CAPPUN_2. Check the wording of each of these questions at the National Opinion Research Center's (NORC) GSS Data Explorer website: GSS Data Explorer | NORC at the University of Chicago (https://gssdataexplorer.norc.gss). "Search Data" and then enter each variable name as a keyword. How well do each of these questions meet the guidelines for writing survey questions?

2. Now generate cross-tabulations to show the relationship between each of these variables, treated as independent variables, and support for capital punishment. A cross-tabulation can be used to display the distribution of responses on the dependent variable for each category of the independent variable. For this purpose, you should substitute several slightly different versions of the variables you just reviewed. From the menu, select Analyze/Descriptive Statistics/

   Crosstabs:

   Rows: **CAPPUN_2**

   Columns: **SEX, PARTYID4, GOVAIDBLK3, FUND2020, OWNGUN_2**

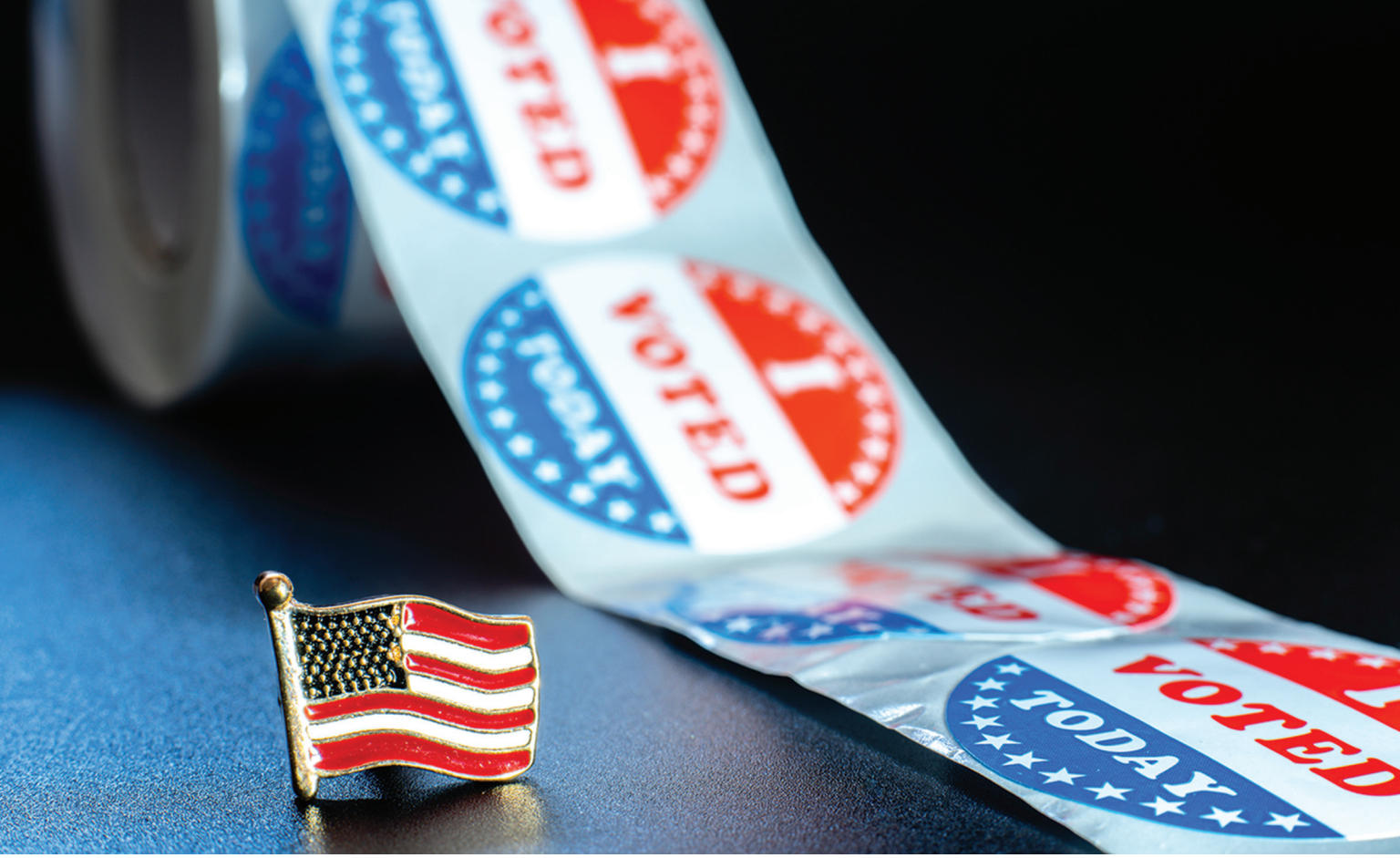
   Cells: **column percentages**

   (If you have had a statistics course, you will also want to request the chi-square statistic for each of these tables.)

   Describe the relationship you have found in the tables, noting the difference in the distribution of the dependent (row) variable—support for capital punishment—between the categories of each of the independent (column) variables.

3. Summarize your findings. What attitudes and characteristics are associated strongly with support for the death penalty?

4. What other hypotheses would you like to test? What questions should be added to the survey to help explain support for the death penalty?



©iStockphoto.com/Joaquin Corbalan

9 QUANTITATIVE DATA ANALYSIS

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

9.1 Explain the use of descriptive statistics

9.2 Describe the steps of coding and data cleaning that are required to prepare for data analysis

9.3 Compare the instances when it is appropriate to use graphs to represent statistics to the instances when it is appropriate to use frequency distributions to represent statistics

9.4 Defend the inclusion of both measures of central tendency and measures of variation in data analysis

9.5 Describe the ways in which cross-tabular analyses can be used to examine the relationship between variables

9.6 Summarize the meaning of the correlation coefficient in a regression analysis

9.7 Describe why it is so easy (knowingly or unknowingly) to present misleading information in summary statistics

Few presidential elections in the history of the United States have been more consequential than the 2020 contest. Record numbers from both parties voted, and the outcome dramatically shifted political leadership, even though it did not end the process of political polarization that preceded it. The ongoing pandemic complicated social science research about the election, but as with many businesses, much data collection shifted online, and so many surveys and other research projects continued, albeit in a somewhat different form. I will use the data collected online in the 2020 General Social Survey to illustrate the use of statistics in this chapter. My examples will focus on explaining the likelihood of voting in the 2020 presidential election.

This chapter introduces several common statistics used in social research and highlights the factors that must be considered when using and interpreting statistics. Think of it as a review of fundamental social statistics, if you have already studied them, or as an introductory overview, if you have not. Two preliminary sections lay the foundation for studying statistics. In the first, I discuss the role of statistics in the research process, returning to themes and techniques with which you are already familiar. In the second preliminary section, I outline the process of preparing data for statistical analysis. In the rest of the chapter, I explain how to describe the distribution of single variables and the relationship between variables. In a final section, I address ethical issues related to data analysis.

Data analysis is an integral component of research methods, and it's important that any proposal for quantitative research include a plan for the analysis that will follow data collection (Chapter 11 shifts the focus to analysis of qualitative data). You must anticipate your data analysis needs if you expect your research design to secure the requisite data. This chapter will have been successful if it encourages you to use statistics responsibly, to evaluate statistics critically, and to seek opportunities for extending your statistical knowledge.

INTRODUCING STATISTICS

Statistics play a key role in achieving valid research results—in measurement, causal validity, and generalizability. Some statistics are useful primarily to describe the results of measuring single variables and to construct and evaluate multi-item scales. These statistics include frequency distributions, graphs, measures of central tendency and variation, and reliability tests. Other statistics are useful primarily in achieving causal validity, by helping us describe the association between variables and control for or otherwise account for other variables. This chapter introduces cross-tabulation as a technique for measuring association and controlling other variables. All such statistics are termed **descriptive statistics** because they are used to describe the distribution of and relationship between variables.

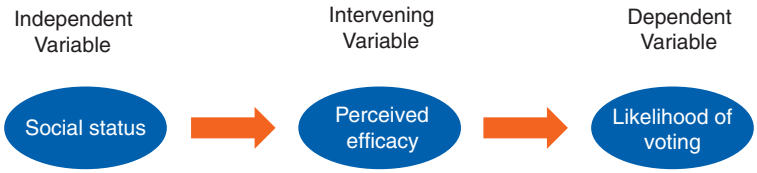
You have already learned in Chapter 5 that it is possible to estimate the degree of confidence that can be placed in a statistical generalization from a sample to the population from which the sample was selected. The statistics used in making these estimates are termed *inferential statistics*. In this chapter, I introduce the use of inferential statistics for testing hypotheses involving sample data.

Social theory and the results of prior research should guide our statistical choices, as they guide the choice of other research methods. There are so many different statistics and so many ways for them to be used in data analysis that even the best statistician can be lost in a sea of numbers if they do not use prior research and theorizing to develop a coherent analysis plan. It is also important to choose statistics that are appropriate to the level of measurement of the variables to be analyzed. As you learned in Chapter 4, numbers used to represent the values of variables may not signify different quantities, meaning that many statistical techniques will be inapplicable for some variables.

Case Study: The Likelihood of Voting

In this chapter, I use for examples data from the 2020 General Social Survey (GSS) on voting and the variables associated with it (Smith et al. 2017), and I will focus on a research question about political participation: What influences the likelihood of voting? Prior research on voting in both national and local settings provides a great deal of support for one hypothesis: The likelihood of voting increases with social status (Manza, Brooks, and Sauder 2005:208; Milbrath and Goel 1977:92–95; Salisbury 1975:326; Verba and Nie 1972:892). Research suggests that social status influences the likelihood of voting through the intervening variable of perceived political efficacy or the feeling that one's vote matters (see Figure 9.1).

FIGURE 9.1 ■ Causal Model of Likelihood of Voting

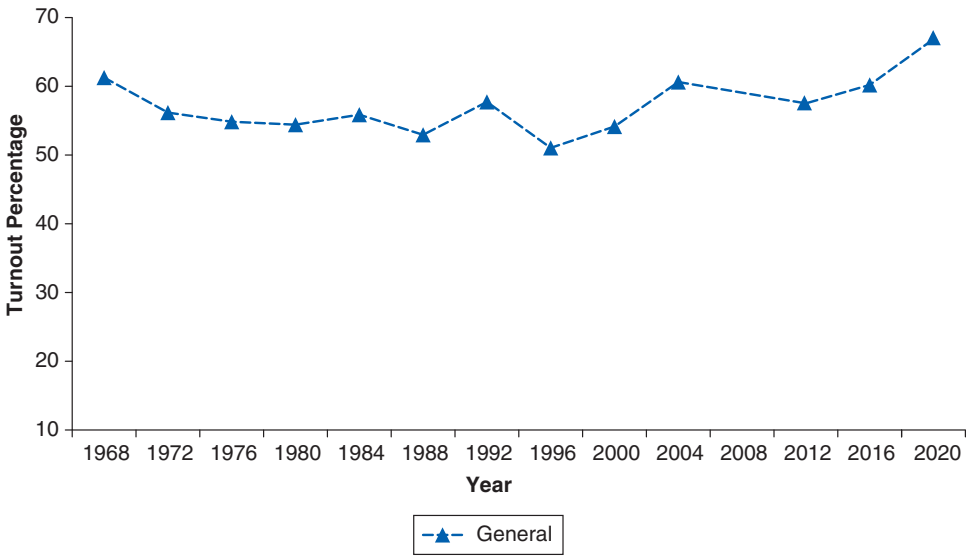


Some research findings on political participation are inconsistent with the social status–voting hypothesis. For example, African Americans participate in politics at higher rates than

do white Americans of similar social status—at least when there is an African American candidate for whom to vote (Manza, Brooks, and Sauder 2005:209; and Verba Nie 1972; Verba Nie and Kim 1978). This discrepant finding suggests the impact of social status on voting and other forms of political participation varies with the social characteristics of potential participants.

The rate of voting of 66.8% of the total citizen voting-age population in the 2020 presidential election exceeded the most recent high of 63.6% in the election of Barack Obama in 2008 (see Figure 9.2) (Gans 2008; Liptak 2012; Wilson 2017). Participation among African Americans at 62.6% dropped somewhat from the preceding high of 66.2% in the 2012 reelection of Barack Obama when it had for the first time, surpassed the participation rate for white non-Hispanic Americans. Turnout increased sharply for Asian Americans to 59.7% and also rose among Hispanic Americans to 53.7%.

FIGURE 9.2 ■ Voting in Presidential Elections, 1980–2020



Sources: Bureau of the Census, Voting and Registration in the Election of November 2020 (percentage of citizen population). Historical Reported Voting Rates (census.gov)

If we are guided by prior research, a test of the hypothesis that the likelihood of voting increases with social status should also account for political efficacy and some social characteristics, such as race. We can find indicators for each of these variables, except political efficacy, in the 2020 GSS (see Table 9.1). We will substitute the variable *interpersonal trust* for political efficacy. I will use these and other GSS2020 variables to illustrate statistics throughout this chapter. You can replicate my analysis with the GSS2020x data set posted on the study site for this book at www.edge.sagepub.com/schutt10e.

TABLE 9.1 ■ List of GSS 2020 Variables Used in Chapter

Variable[a]	SPSS Variable Name	Description
Family income	INCOME16_2	Family income (in categories)
	INCOME4	Family income (in 4 categories)
Respondent's income	RINCOMEX	Respondent's income (in approximate dollars)
Region	REGION4	East, Midwest, South, West
Education	EDUC_2	Highest year of school completed
	EDUC6	Years of education completed (6 categories)
	EDUC4	Years of education (4 categories)
Age	AGE_2	Years old
	AGE7	Years old (in decades)
Gender	FEMALE	Male, female
Confidence in institutions	CONARMY_2 TO CONTV_2	Level of confidence (a great deal, only some, hardly any)
Attitude toward abortion	ABANY_2 TO ABSINGLE_2	Agreement with abortion decisions, given conditions (yes, no)
Marital status	MARITAL_2	Married, never married, widowed, divorced
Race	RACE2020	Respondent's race (white, Black, other)
Likelihood of voting	LIKELYVOTE3	Likelihood of voting in 2020 presidential election (extremely, very or moderately, slightly, or not at all)
Approval of President Trump	TRUMP2020	Strongly approve or approve, strongly disapprove or disapprove
Interpersonal trust	TRUST_2	Believe other people can be trusted (yes, no)

a. Some variables recoded.

Source: General Social Survey, National Opinion Research Center 2020.

PREPARING FOR DATA ANALYSIS

My analysis of voting in this chapter is an example of secondary data analysis, which you will learn about in Chapter 14. Using secondary data in this way has a major advantage: The researcher doesn't have to secure the funds and spend the time required to collect their own data. But there are also disadvantages: If you did not design the study yourself, it is unlikely all the variables you think should have been included were included and were measured in the way you prefer (for example, I had to substitute a measure of interpersonal trust for a measure of political efficacy). In addition, the sample may not represent just the population in which you are interested, and the study design may be only partially appropriate to your research question. For example, because it is a survey of individuals, the GSS2020 lacks measures of political context (such as the dominant party in an area). Because the survey sample is selected only from the United States and because the questions concern just one presidential election, we will not be able to directly address the larger issues of political context that are represented in cross-national and longitudinal research (for more on cross-national and longitudinal research, see Verba et al. 1978).

If you conduct your own survey or experiment, your data must be prepared in a format suitable for computer entry; termed **data entry**. Several options are available. Questionnaires can be designed so responses are recorded directly in a computer database, either by the interviewer or by the respondent in web-based surveys (see Figure 9.3). However, when responses are recorded on paper-and-pencil questionnaires, they must be entered into a database as a separate process, either through keyboard entry or through scanning a properly prepared form. **Coding** of all responses should be done before data entry by assigning each response a unique numerical value. If the survey is distributed through the Internet or completed directly on a computer, database software can be programmed to recognize the response codes and record them as numbers when the survey responses are entered.

FIGURE 9.3 ■ Questions from a Web-Based Survey

Q29 During the **last 7 days**, how much money did you or your household spend on prepared meals, including eating out, fast food, and carry out or delivered meals? Please include money spent in cafeterias at work or at school or on vending machines. Please do not include money you have already told us about in item Q28(above). *Enter amount.*

Q30 How confident are you that your household will be able to afford the kinds of food you need for the **next four weeks**? *Select only one answer.*

- ○ Not at all confident
- ○ Somewhat confident
- ○ Moderately confident
- ○ Very confident

Source: U.S. Bureau of the Census. 2020 Household Pulse Survey.

Whatever data entry method is used, the data must be checked carefully for errors—a process called **data cleaning**. Most data entry programs that enter survey responses into a database can be programmed to check each response before it is saved and ensure it represents a valid response for that variable. When an invalid entry occurs or no response is entered for a required question, the program prompts the respondent to try again. This process reduces sharply the possibility of data entry errors.

After a database with survey responses is created, a computer program must be written to "define the data." A data definition program identifies the variables that are coded in each column or range of columns, attaches meaningful labels to the codes, and distinguishes values

representing missing data. The procedures for doing so vary with the specific statistical package used. I used the Statistical Package for the Social Sciences (SPSS) for the analysis in this chapter; you will find examples of SPSS commands for defining and analyzing data in Appendix D (on the study site at edge.sagepub.com/schutt10e). More information on using SPSS is contained in SPSS manuals and in the SAGE Publications volume *Using IBM® SPSS® Statistics for Research Methods and Social Science Statistics*, 7th edition, by William E. Wagner III (2019).

DISPLAYING UNIVARIATE DISTRIBUTIONS

The first step in data analysis is usually to display the variation in each variable of interest. For many descriptive purposes, the analysis may go no further. Graphs and frequency distributions are the two most popular approaches; both allow the analyst to display the distribution of cases across the categories of a variable. Graphs have the advantage of providing a picture that is easier to comprehend, although frequency distributions are preferable when exact numbers of cases having particular values must be reported and when many distributions must be displayed in a compact form.

Whichever type of display is used, the primary concern of the data analyst is to display accurately the distribution's shape—that is, to show how cases are distributed across the values of the variable. Three features of shape are important: **central tendency**, **variability**, and **skewness** (lack of symmetry). All three features can be represented in a graph or in a frequency distribution.

These features of a distribution's shape can be interpreted in several different ways, and they are not all appropriate for describing every variable. In fact, all three features of a distribution can be distorted if graphs, frequency distributions, or summary statistics are used inappropriately.

A variable's level of measurement is the most important determinant of the appropriateness of particular statistics. For example, we cannot talk about the skewness (lack of symmetry) of a variable measured at the nominal level. If the values of a variable cannot be ordered from lowest or highest—if the ordering of the values is arbitrary—we cannot say that the distribution is not symmetric because we could just reorder the values to make the distribution more (or less) symmetric. Some measures of central tendency and variability are also inappropriate for variables measured at the nominal level.

The distinction between variables measured at the ordinal level and those measured at the interval or ratio level should also be considered when selecting statistics for use, but social researchers differ in just how much importance they attach to this distinction. Many social researchers think of ordinal variables as imperfectly measured interval-level variables and believe that, in most circumstances, statistics developed for interval-level variables also provide useful summaries for ordinal variables. Other social researchers believe that variation in ordinal variables will often be distorted by statistics that assume an interval level of measurement. We will touch on some of the details in the following sections on particular statistical techniques.

We will now examine graphs and frequency distributions that illustrate these three features of shape. Summary statistics used to measure specific aspects of central tendency and variability are presented in a separate section. There is a summary statistic for the measurement of skewness, but it is used only rarely in published research reports and will not be presented here.

Graphs

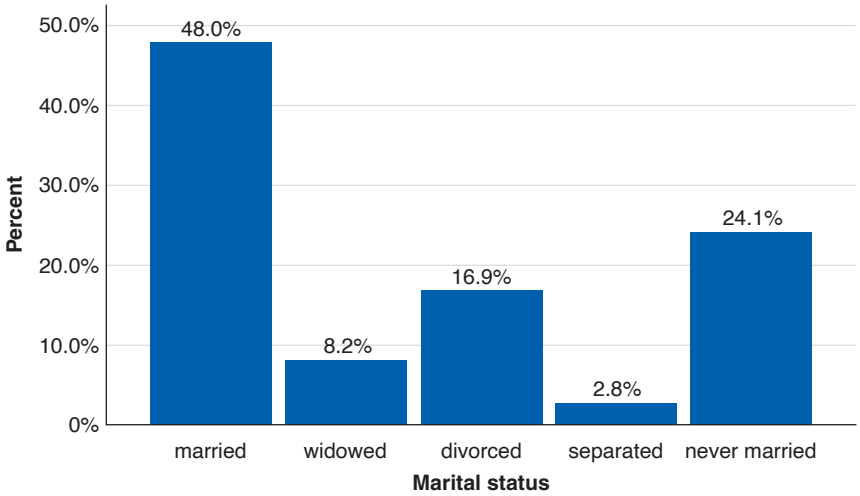
A picture often is worth some unmeasurable quantity of words. Even for the uninitiated, graphs can be easy to read, and they highlight a distribution's shape. They are useful particularly for exploring data because they show the full range of variation and identify data anomalies that

might be in need of further study. And good, professional-looking graphs can now be produced relatively easily with software available for personal computers. There are many types of graphs, but the most common and most useful are bar charts, histograms, and frequency polygons. Each has two axes, the vertical axis (the *y*-axis) and the horizontal axis (the *x*-axis), and labels to identify the variables and the values, with tick marks showing where each indicated value falls along the axis.

Bar Chart

A **bar chart** contains solid bars separated by spaces. It is a good tool for displaying the distribution of variables measured at the nominal level because there is, in effect, a gap between each of the categories. The bar chart of marital status in Figure 9.4 indicates almost half of adult Americans were married at the time of the survey. Smaller percentages were divorced, separated, or widowed, and almost one quarter had never married. The most common value in the distribution is married, so this would be the distribution's central tendency. There is a moderate amount of variability in the distribution because the half that are not married are spread across the categories of widowed, divorced, separated, and never married. Because marital status is not a quantitative variable, the order in which the categories are presented is arbitrary, and so skewness is not relevant.

FIGURE 9.4 ■ Bar Chart of Marital Status

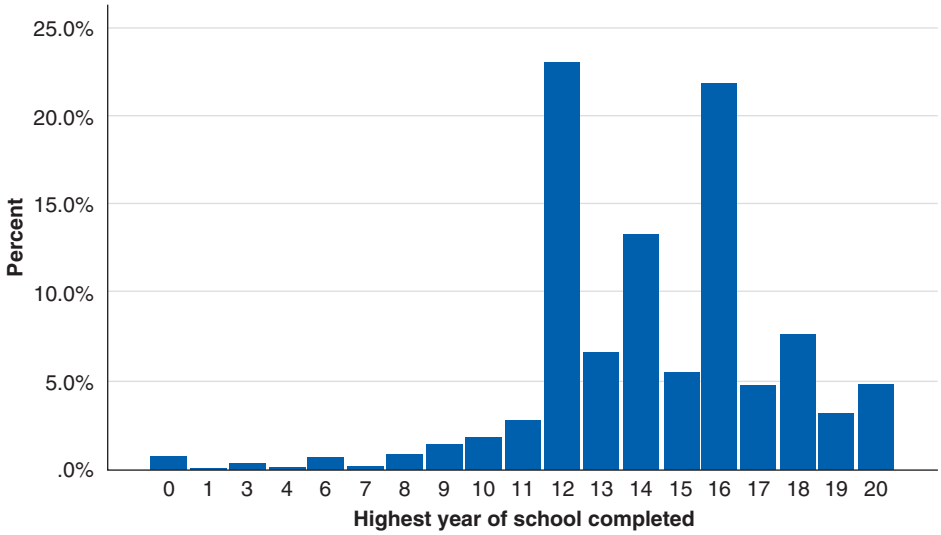


Source: General Social Survey, National Opinion Research Center 2020.

Histograms

Histograms, in which the bars are adjacent, are used to display the distribution of quantitative variables that vary along a continuum that has no necessary gaps. Figure 9.5 shows a histogram of years of education from the 2020 GSS data. The distribution has a clump of cases at 12 years—about one third of the total—and another at 16 years, corresponding, respectively, to completion of high school and of college. The distribution is negatively skewed because the percentage of cases tapers off to the low end much more quickly, with a long tail, compared with the shape of the distribution toward the high end.

FIGURE 9.5 ■ Histogram of Years of Education

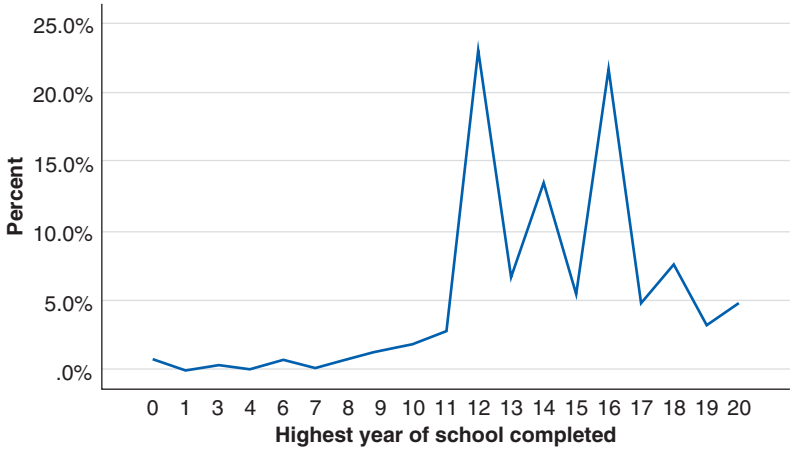


Source: General Social Survey, National Opinion Research Center 2020.

Frequency Polygon

In a **frequency polygon**, a continuous line connects the points representing the number or percentage of cases with each value. The frequency polygon is an alternative to the histogram when the distribution of a quantitative variable must be displayed; this alternative is particularly useful when the variable has a wide range of values. It is easy to see in the frequency polygon of years of education in Figure 9.6 that the most common value is 12 years, high school completion, and that this value also seems to be at the center of the distribution. There is moderate variability in the distribution, with many cases having more than 12 years of education and almost one third having completed at least four years of college (16 years). The distribution is highly skewed in the negative direction, with few respondents reporting fewer than 10 years of education.

FIGURE 9.6 ■ Frequency Polygon of Years of Education

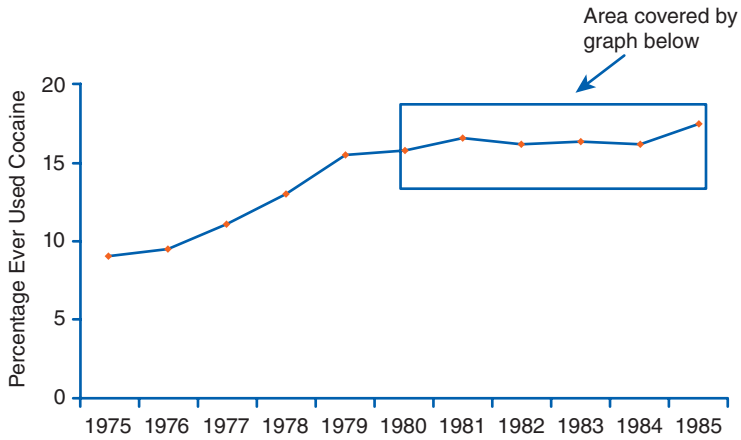


Source: General Social Survey, National Opinion Research Center 2020.

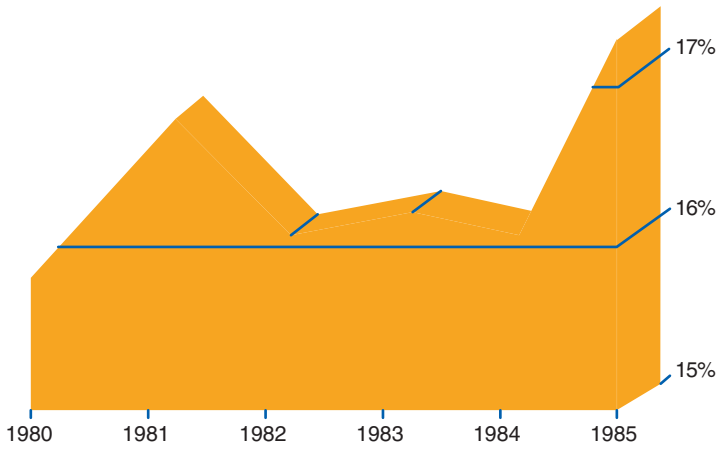
Misuse of Graphs

If graphs are misused, they can distort rather than display the shape of a distribution. Compare, for example, the two graphs in Figure 9.7. The first graph shows that high school seniors reported relatively stable rates of lifetime use of cocaine between 1980 and 1985. The second graph, using exactly the same numbers, appeared in a 1986 *Newsweek* article on the coke plague (Orcutt and Turner 1993). Looking at this graph, you would think the rate of cocaine usage among high school seniors had increased dramatically during this period. But, in fact, the difference between the two graphs is due simply to changes in how the graphs are drawn. In the "plague graph" the percentage scale on the vertical axis begins at 15 rather than at 0, making what was about a 1-percentage point increase look very big indeed. In addition, omission from the plague graph of the more rapid increase in reported usage between 1975 and 1980 makes it look as if the tiny increase in 1985 were a unique and thus more newsworthy crisis.

FIGURE 9.7 ■ Two Graphs of Cocaine Usage



A. University of Michigan Institute for Social Research
Time Series for Lifetime Prevalence of Cocaine Use



B. Final Stages of Construction

Source: Adapted from Orcutt and Turner (1993). Copyright 1993 by the Society for the Study of Social Problems. Reprinted by permission.

Adherence to several guidelines (Tufte 1983; Wallgren et al. 1996) will help you spot these problems and avoid them in your own work:

- The difference between bars can be exaggerated by cutting off the bottom of the vertical axis and displaying less than the full height of the bars. Instead, begin the graph of a quantitative variable at 0 on both axes. It may be reasonable, at times, to violate this guideline, as when an age distribution is presented for a sample of adults, but in this case, be sure to mark the break clearly on the axis.
- Bars of unequal width, including pictures instead of bars, can make particular values look as if they carry more weight than their frequency warrants. Always use bars of equal width.
- Either shortening or lengthening the vertical axis will obscure or accentuate the differences in the number of cases between values. The two axes usually should be of approximately equal length.
- Avoid chart junk that can confuse the reader and obscure the distribution's shape (a lot of verbiage or umpteen marks, lines, lots of cross-hatching, etc.).

Frequency Distributions

A **frequency distribution** displays the number of cases, percentage (the relative frequencies) of cases, or both, corresponding to each of a variable's values or group of values. The components of the frequency distribution should be clearly labeled, with a title, a stub (labels for the values of the variable), a caption (identifying whether the distribution includes frequencies, percentages, or both), and perhaps the number of missing cases. If percentages rather than frequencies are presented (sometimes both are included), the total number of cases in the distribution (the **base number *N***) should be indicated (see Table 9.2).

TABLE 9.2 ■ Frequency Distribution of Likelihood of Voting in the 2020 Presidential Election

Value	Frequency	Valid Percentage
Extremely likely to vote	1,250	68.6%
Very likely to vote	186	10.8
Moderately likely to vote	90	5.2
Slightly likely to vote	51	3.0
Not likely at all to vote	145	8.4
No answer	101	
Total %		100.0%
N	1,823	(1,722)

Source: General Social Survey, National Opinion Research Center 2020.

Ungrouped Data

Constructing and reading frequency distributions for variables with few values is not difficult. The frequency distribution of voting in Table 9.2, for example, shows that 68.6% of the respondents rated themselves as extremely likely to vote, while another 16% said they were very likely (10.8) or moderately likely (5.2) to vote. Just 3% estimated they were only slightly likely to vote, while 8.4% were not likely to vote. The total number of respondents to this question was 1,722, with another 101 not answering the question.

If you compare Table 9.2 with Figure 9.6, you can see that a frequency distribution can provide more precise information than a graph about the number and percentage of cases in a variable's categories. Often, however, it is easier to see the shape of a distribution when it is graphed. When the goal of a presentation is to convey a general sense of a variable's distribution, particularly when the presentation is to an audience not trained in statistics, a graph is often the best choice.

Grouped Data

Many frequency distributions (and graphs) require grouping of some values after the data are collected. There are two reasons for grouping:

1. There are more than 15–20 values to begin with, a number too large to be displayed in an easily readable table.
2. The distribution of the variable will be clearer or more meaningful if some of the values are combined.

Inspection of Table 9.3 should clarify these reasons. In the first distribution, which is only a portion of the entire ungrouped GSS age distribution, it is very difficult to discern any shape, much less the central tendency. In the second distribution, age is grouped in the familiar 10-year intervals and the distribution's shape is immediately clear.

TABLE 9.3 ■ Ungrouped Versus Grouped Frequency Distributions

Ungrouped		Grouped	
Age	**Percentage**	**Age**	**Percentage**
		20–29	10.5%
20	.45	30–39	17.4
21	.7	40–49	16.3
22	.4	50–59	17.2
23	1.0	60–69	18.7
24	.7	70–79	14.8
25	.9	80 or older	5.1
26	1.0		100.0%
27	1.5		(1823)

Ungrouped		Grouped	
Age	**Percentage**	**Age**	**Percentage**
28	1.4		
29	1.6		
30	1.6		
31	2.1		
32	1.5		
33	1.7		
34	1.8		
35	1.6		
36	1.9		
37	1.8		
38	1.5		
...	...		

Source: General Social Survey National Opinion Research Center 2020.

Once we decide to group values or categories, we must be sure that in doing so we do not distort the distribution. Adhering to the following guidelines for combining values in a frequency distribution will prevent many problems:

- Categories should be logically defensible and preserve the distribution's shape.
- Categories should be mutually exclusive and exhaustive, so every case should be classifiable in one and only one category.

Violating these two guidelines is easier than you might think. If you were to group all the ages above 59 together as 60 or higher, it would create the appearance of a bulge at the high end of the age distribution, with 38.6% of the cases. Combining other categories so they include a wide range of values could create the same type of misleading impression. In some cases, however, the most logically defensible categories will vary in size. A good example would be grouping years of education to progress through grade school, high school, college, and graduate work. Such a grouping captures the most meaningful distinctions in the educational distribution and preserves the information that would be important for many analyses (see Table 9.4).

It is also easy to imagine how the requirement that categories be mutually exclusive can be violated. You sometimes see frequency distributions or categories in questionnaires that use such overlapping age categories as 20–30, 30–40, and so on, instead of mutually exclusive categories, such as those in Table 9.3. The problem when the categories are not mutually exclusive is we then can't tell which category to place someone in who is age 30, 40, and so on.

TABLE 9.4 ■ Years of Education Completed

Years of Education	Percentage
Less than 8	2.3%
8–11	7.0%
12	23.1%
13–15	25.4%
16	21.8%
17 or more	20.5%
	100.0 (1,813)

Source: General Social Survey, National Opinion Research Center 2020.

Combined and Compressed Distributions

In a **combined frequency display**, the distributions for a set of conceptually similar variables having the same response categories are presented together. Table 9.5 is a combined display reporting the frequency distributions in percentage form for 13 variables, which indicate GSS

TABLE 9.5 ■ Confidence in Institutions

Confidence in...	A Great Deal	(%) Only Some	(%) Hardly Any	(%) Total	*N*
Military	49.8%	37.6	12.6	100.0%	1,238
Scientific community	44.7%	46.5	8.8	100.0%	1,234
Medicine	36.7%	51.8	11.4	100.0%	1,241
U.S. Supreme Court	24.1%	55.8	20.1	100.0%	1,236
Education	21.1%	56.8	22.1	100.0%	1,239
Banks and financial institutions	18.2%	57.1	24.7	100.0%	1,239
Major companies	16.0%	60.7	23.3	100.0%	1,234
Organized religion	15.5%	51.1	33.4	100.0%	1,236
Executive branch of federal government	13.0%	33.4	53.6	100.0%	1,232
Press	12.5%	39.8	47.7	100.0%	1,235
Organized labor	9.7%	63.5	26.8	100.0%	1,223
Television	8.4%	48.1	43.5	100.0%	1,236
Congress	4.9%	41.6%	53.5%	100.0%	1.232

Source: General Social Survey, National Opinion Research Center 2020.

respondents' level of confidence in U.S. institutions. The different variables are identified in the leftmost column, and their values are labeled along the top. By looking at the table, you can see quickly that confidence is greatest in the military, the scientific community, and medicine, while the Supreme Court, educational institutions, banks and financial institutions, and major companies are regarded with somewhat less confidence. Television and the U.S. Congress elicit the least confidence. Note that I have ordered the specific variables in decreasing order of greatest confidence to make it easier to see the pattern. The number of cases on which the distributions are based is included for each variable.

Compressed frequency displays can also be used to present cross-tabular data and summary statistics more efficiently, by eliminating unnecessary percentages (such as those corresponding to the second value of a dichotomous variable) and by reducing the need for repetitive labels. Table 9.6 presents a compressed display of agreement that abortion should be allowed given particular conditions. Note this display presents (in parentheses) the number of cases on which the percentages are based.

TABLE 9.6 ■ Agreement With Allowing Abortion, Given Different Conditions

Statement	% Yes	*N*
Woman's health seriously endangered	88.1%	(1,175)
Pregnant as result of rape	81.2%	(1,118)
Strong chance of serious defect	73.9%	(1,178)
Low income; can't afford more children	55.9%	(1,180)
Married and wants no more children	54.6%	(1,185)
Abortion if woman wants for any reason	52.9%	(1,186)
Not married	51.5%	(1,182)

Source: General Social Survey, National Opinion Research Center 2020.

Combined and compressed statistical displays facilitate the presentation of a large amount of data in a relatively small space. They should be used with caution, however, because they may baffle people who are not used to them.

SUMMARIZING UNIVARIATE DISTRIBUTIONS

Summary statistics focus attention on particular aspects of a distribution and facilitate comparison between distributions. For example, if your purpose were to report variation in income by state in a form that is easy for most audiences to understand, you would usually be better off presenting average incomes; many people would find it difficult to make sense of a display containing 50 frequency distributions, although they could readily comprehend a long list of average incomes. A display of average incomes would also be preferable to multiple frequency distributions if your only purpose were to provide a general idea of income differences between states.

RESEARCH IN THE NEWS

We Did Not Suffer Equally

When the pandemic shut down activities around the world, the phrase "we're all in this together" was often heard. Yet as the pandemic persisted, many higher-income employees were able to shift to working online from the safety of their homes, while many lower-income employees either lost their jobs as their businesses, like hotels and restaurants and many factories, remained closed, or employees had to risk infection by continuing their work in essential industries, like hospitals and food preparation. Life expectancy fell by almost a year for white Americans and almost three years for Black Americans. Women were more likely to take on childcare and educational activities at home, which prevented them from maintaining their previous occupations. Use of food banks increased. Overall, preexisting disparities increased and the sense of union in the face of a common threat diminished.

For Further Thought?

1. Would you expect the overall level of economic harm during the pandemic to form a somewhat normal curve, a bimodal distribution, or some other shape? Explain your reasoning.
2. What third variable do you think would be most likely to explain differences in the pandemic's economic effects by race or gender?

News source: Serkez, Yaryna. 2021. "We Did Not Suffer Equally: The Pandemic Has Been a Story of Two Lines—Haves and Have-Nots—Moving in Different Directions." *The New York Times*, March 14:SR9.

Of course, representing a distribution in one number loses information about other aspects of the distribution's shape and so creates the possibility of obscuring important information. If you need to inform a discussion about differences in income inequality between states, for example, measures of central tendency and variability would miss the point entirely. You would either have to present the 50 frequency distributions or use some special statistics that represent the unevenness of a distribution. For this reason, analysts who report summary measures of central tendency usually also report a summary measure of variability and sometimes several measures of central tendency, variability, or other statistics.

Measures of Central Tendency

Central tendency is usually summarized with one of three statistics: the mode, the median, or the mean. For a particular application, one of these statistics may be preferable, but each has a role to play in data analysis. To choose an appropriate measure of central tendency, the analyst must consider a variable's level of measurement, the skewness of a quantitative variable's distribution, and the purpose for which the statistic is used. In addition, the analyst's personal experiences and preferences inevitably will play a role.

Mode

The **mode** is the most frequent value in a distribution. It is also termed the **probability average** because, being the most frequent value, it is the most probable. For example, if you were to pick a case at random from the distribution of likelihood of voting in 2020 (refer back to Table 9.2), the probability of the case being extremely likely to vote would be .686 out of 1, or 68.6%—the most probable value in the distribution.

The mode is used much less often than the other two measures of central tendency because it can so easily give a misleading impression of a distribution's central tendency. One problem with the mode occurs when a distribution is **bimodal** in contrast to being **unimodal**. A bimodal (or trimodal, etc.) distribution has two or more categories with an equal number of cases and with more cases than the other categories. There is no single mode in a bimodal distribution. Imagine that a particular distribution has two categories, each having just about the same number of cases (and these are the two most frequent categories). Strictly speaking, the mode would be the one with more cases, even though the other frequent category had only slightly fewer cases. Another potential problem with the mode is that it might happen to fall far from the main clustering of cases in a distribution. It would be misleading in most circumstances to say simply that the variable's central tendency was whatever the modal value was.

Nevertheless, there are occasions when the mode is very appropriate. Most important, the mode is the only measure of central tendency that can be used to characterize the central tendency of variables measured at the nominal level. We can't say much more about the central tendency of the distribution of marital status in Figure 9.4 than that the most common value is married. The mode also is often referred to in descriptions of the shape of a distribution. The terms *unimodal* and *bimodal* appear frequently, as do descriptive statements such as "The typical [most probable] respondent was in her 30s." Of course, when the issue is what the most probable value is, the mode is the appropriate statistic. Which ethnic group is most common in a school? The mode provides the answer.

Median

The **median** is the position average, or the point that divides the distribution in half (the 50th percentile). The median is inappropriate for variables measured at the nominal level because their values cannot be put in order, and so there is no meaningful middle position. To determine the median, we simply array a distribution's values in numerical order and find the value of the case that has an equal number of cases above and below it. If the median point falls between two cases (which happens if the distribution has an even number of cases), the median is defined as the average of the two middle values and is computed by adding the values of the two middle cases and dividing by 2.

The median in a frequency distribution is determined by identifying the value corresponding to a cumulative percentage of 50. Starting at the top of the years of education distribution in Table 9.4, for example, and adding up the percentages, we find we have reached 42% in the 12 years category and then 67.8% in the 13–15 years category. The median is therefore 13–15.

With most variables, it is preferable to compute the median from ungrouped data because that method results in an exact value for the median rather than an interval. In the grouped age distribution in Table 9.3, for example, the median is in the 50s interval. But if we determine the median by adding up the percentages in succession for all the individual ages in the ungrouped data, we would find the exact value of the median is 53.

Mean

The **mean**, or arithmetic average, considers the values of each case in a distribution—it is a weighted average. The mean is computed by adding up the value of all the cases and dividing by the total number of cases, thereby accounting for the value of each case in the distribution:

$$\text{Mean} = \text{Sum of value of cases}/\text{Number of cases}$$

In algebraic notation, the equation is $\overline{Y} = \sum Y_i / N$. For example, to calculate the mean of eight (hypothetical) cases, we add the values of all the cases (ΣYi) and divide by the number of cases (N):

$$(28 + 117 + 42 + 10 + 77 + 51 + 64 + 55)/8 = 444/8 = 55.5$$

Because computing the mean requires adding the values of the cases, it makes sense to compute a mean only if the values of the cases can be treated as actual quantities—that is, if they reflect an interval or ratio level of measurement or if they are ordinal and we assume that ordinal measures can be treated as if they are measures at the interval level. It would make no sense to calculate the mean religion. For example, imagine a group of four people in which two were Protestants, one Catholic, and one Jewish. To calculate the mean, you would need to solve the equation (Protestant + Protestant + Catholic + Jewish) + 4 = ? Even if you decide that Protestant = 1, Catholic = 2, and Jewish = 3 for data entry purposes, it still doesn't make sense to add these numbers because they don't represent quantities of religion.

Median or Mean?

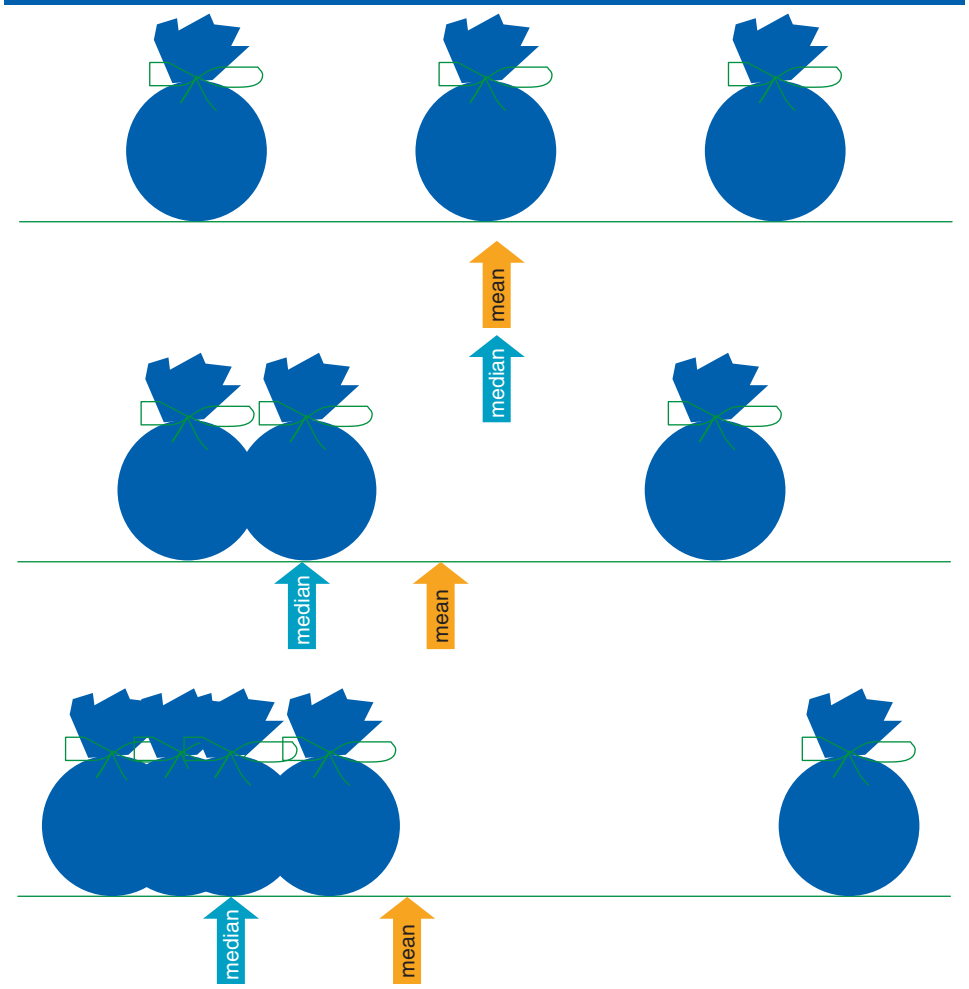
Both the median and the mean are used to summarize the central tendency of quantitative variables, but their suitability for a particular application must be carefully assessed. The key issues to be considered in this assessment are the variable's level of measurement, the shape of its distribution, and the purpose of the statistical summary. Consideration of these issues will sometimes result in a decision to use both the median and the mean and will sometimes result in neither measure being seen as preferable. But in many other situations, the choice between the mean and median will be clear-cut as soon as the researcher takes the time to consider these three issues.

Level of measurement is a key concern because to calculate the mean, we must add the values of all the cases—a procedure that assumes the variable is measured at the interval or ratio level. So even though we know that coding *Agree* as 2 and *Disagree* as 3 does not really mean that *Disagree* is 1 unit more of disagreement than *Agree*, the mean assumes this evaluation to be true. Because calculation of the median requires only that we order the values of cases, we do not have to make this assumption. Technically speaking, then, the mean is simply an inappropriate statistic for variables measured at the ordinal level (and you already know that it is completely meaningless for variables measured at the nominal level). In practice, however, many social researchers use the mean to describe the central tendency of variables measured at the ordinal level, for the reasons outlined earlier.

The shape of a variable's distribution should also be considered when deciding whether to use the median or mean. When a distribution is perfectly symmetric, so the distribution of values below the median is a mirror image of the distribution of values above the median, the mean and median will be the same. But the values of the mean and median are affected differently by skewness, or the presence of cases with extreme values on one side of the distribution but not the other side. Because the median accounts for only the number of cases above and below the median point, not the value of these cases, it is not affected in any way by extreme values. Because the mean is based on adding the value of all the cases, it will be pulled in the direction of exceptionally high (or low) values. When the value of the mean is larger than the value of the median, we know the distribution is skewed in a positive direction, with proportionately more cases with higher than lower values. When the mean is smaller than the median, the distribution is skewed in a negative direction.

This differential impact of skewness on the median and mean is illustrated in Figure 9.8. On the first balance beam, the cases (bags) are spread out equally, and the median and mean are in

FIGURE 9.8 ■ The Mean as a Balance Point



the same location. On the second and third balance beams, the median corresponds to the value of the middle case, but the mean is pulled toward the value of the one case with an extremely high value. The more extreme the high (or low) values, the more the value of the mean will differ from the value of the median.

The single most important influence on the choice of the median or the mean for summarizing the central tendency of quantitative variables should be the purpose of the statistical summary. If the purpose is to report the middle position in one or more distributions, then the median is the appropriate statistic, whether or not the distribution is skewed (see Table 9.7). For example, with respect to the age distribution from the GSS, you could report that half the U.S. population is younger than 53 years old and half the population is older than that. But if the purpose is to show how likely different groups are to have age-related health problems, then the measure of central tendency for these groups should account for people's actual ages, not just the number of people who are older and younger than a particular age. For this purpose, the median would be inappropriate because it would not distinguish between two distributions that have the same median but with different numbers of older people. In one distribution, everyone might be between the ages of 35 and 55, with a median of 49. In another distribution with a median of 49,

TABLE 9.7 ■ Selection of Measures of Central Tendency (MCT)

Level of Measurement	Purpose of MCT	Most Appropriate MCT	Additional Potentially Useful MCT	Definitely Inappropriate MCT
Nominal	Identify most frequent response	Mode	None	Median, mean
Ordinal	Identify middle position	Median	Mean mode	None
Interval, ratio	Identify arithmetic average	Mean	Median, mode	None

half of the cases could have ages above 60. The mean of the second distribution would be higher, reflecting the fact that it has a greater number of older people.

Keep in mind that it is not appropriate to use either the median or the mean as a measure of central tendency for variables measured at the nominal level because, at this level, the different attributes of a variable cannot be ordered as higher or lower (as reflected in Table 9.7). Technically speaking, the mode should be used to measure the central tendency of variables measured at the nominal level (and it can also be used with variables measured at the ordinal, interval, and ratio levels). The median is most suited to measure the central tendency of variables measured at the ordinal level (and it can also be used to measure the central tendency of variables measured at the interval and ratio levels). Finally, the mean is suited to measure the central tendency for variables measured at the interval and ratio levels.

It is not entirely legitimate to represent the central tendency of a variable measured at the ordinal level with the mean: Calculation of the mean requires summing the values of all cases, and at the ordinal level, these values indicate only order, not actual numbers. Nonetheless, many social scientists use the mean with ordinal-level variables and find that this is potentially useful for comparisons between variables and as a first step in more complex statistical analyses.

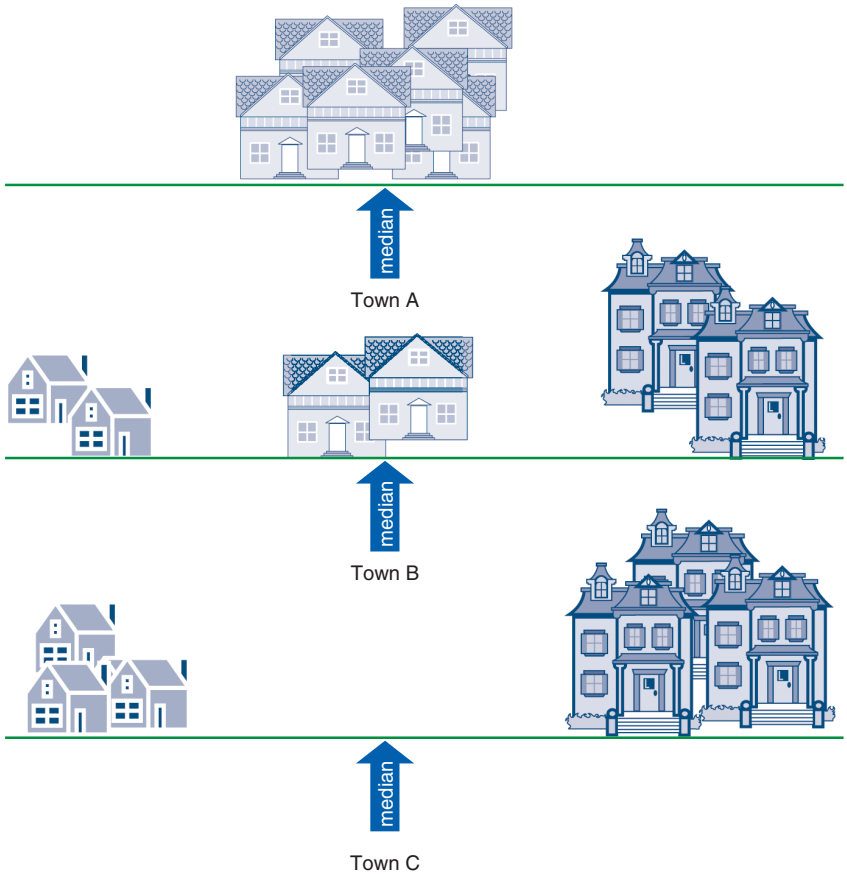
In general, the mean is the most commonly used measure of central tendency for quantitative variables, both because it accounts for the value of all cases in the distribution and because it is the foundation for many more advanced statistics. However, the mean's very popularity results in its use in situations for which it is inappropriate. Keep an eye out for this problem.

Measures of Variation

You already have learned that central tendency is only one aspect of the shape of a distribution—the most important aspect for many purposes but still just a piece of the total picture. A summary of distributions based only on their central tendency can be incomplete, even misleading. For example, three towns might have the same median income but still be very different in their social character because of the shape of their income distributions (as reflected in housing values) As illustrated in Figure 9.9, Town A is a homogeneous middle-class community; Town B is very heterogeneous, with some people living in small homes, some in middle-range homes, and some in very large homes. Town C has a polarized, bimodal income distribution, with mostly very poor and very rich people and few in between. However, all three towns have the same median income.

The way to capture these differences is with statistical measures of variation. Four popular measures of variation are the range, the interquartile range, the variance, and the standard

FIGURE 9.9 ■ Distributions of Housing Values Differing in Variability but Not Central Tendency



deviation (which is the most popular measure of variability). To calculate each of these measures, the variable must be at the interval or ratio level (but many would argue that, like the mean, they can be used with ordinal-level measures, too). Statistical measures of variation are used infrequently with variables measured at the nominal level, so these measures will not be presented here.

It's important to realize that measures of variability are summary statistics that capture only part of what we need to be concerned with about the distribution of a variable. They do not tell us about the extent to which a distribution is skewed, which we've seen is very important for interpreting measures of central tendency. Researchers usually evaluate the skewness of distributions just by eyeballing them.

Range

The **range** is a simple measure of variation, calculated as the highest value in a distribution minus the lowest value:

$$\text{Range} = \text{Highest value} - \text{Lowest value}$$

It often is important to report the range of a distribution to identify the whole range of possible values that might be encountered. However, because the range can be drastically altered by just one exceptionally high or low value (termed an **outlier**), it does not do an adequate job of summarizing the extent of variability in a distribution.

Interquartile Range

A version of the range statistic, the **interquartile range**, avoids the problem created by outliers. **Quartiles** are the points in a distribution corresponding to the first 25% of the cases, the first 50% of the cases, and the first 75% of the cases. You already know how to determine the second quartile, corresponding to the point in the distribution covering half of the cases—it is another name for the median. The first and third quartiles are determined in the same way but by finding the points corresponding to 25% and 75% of the cases, respectively. The interquartile range is the difference between the end of the first quartile and the beginning of the third quartile.

We can use the distribution of age for an example. If you add up the percentages corresponding to each value of age (ungrouped) in Figure 9.9, you'll find that you reach the first quartile (25% of the cases) at the age value of 38. If you were to continue, you would find that age 66 corresponds to the third quartile—the point where you have covered 75% of the cases. So the interquartile range for age, in the GSS 2020 panel data, is 28:

$$\text{Third quartile} - \text{First quartile} = \text{Interquartile range}$$
$$66 - 38 = 28$$

Variance

The **variance** is the average squared deviation of each case from the mean, so it accounts for the amount by which each case differs from the mean. An example of how to calculate the variance, using the following formula, appears in Table 9.8:

$$\sigma^2 = \frac{\sum(Y_i - \overline{Y}_i)}{N}$$

TABLE 9.8 ■ Calculation of the Variance

Case #	Score (Y_i)	$Y_i - \overline{Y}$	$(Y_i - \overline{Y})^2$
1	21	-3.27	10.69
2	30	5.73	32.83
3	15	-9.27	85.93
4	18	-6.27	39.31
5	25	0.73	0.53
6	32	7.73	59.75
7	19	-5.27	27.77
8	21	-3.27	10.69
9	23	-1.27	1.61
10	37	12.73	162.05
11	26	1.73	2.99
Σ			434.15

Symbol key: $\overline{Y}$ = mean; N = number of cases; Σ = sum over all cases; Yi = value of variable Y for case i; √= square root; σ = standard deviation.

Symbol key: $\bar{Y}$ = mean; N = number of cases; Σ = sum over all cases; Yi = value of variable Y for case i; σ^2 = variance.

You can see in the last column of Table 9.8 an example of summing over all cases represented by the Greek letter Σ.

The variance is used in many other statistics, although it is more conventional to measure variability with the closely related standard deviation than with the variance.

Standard Deviation

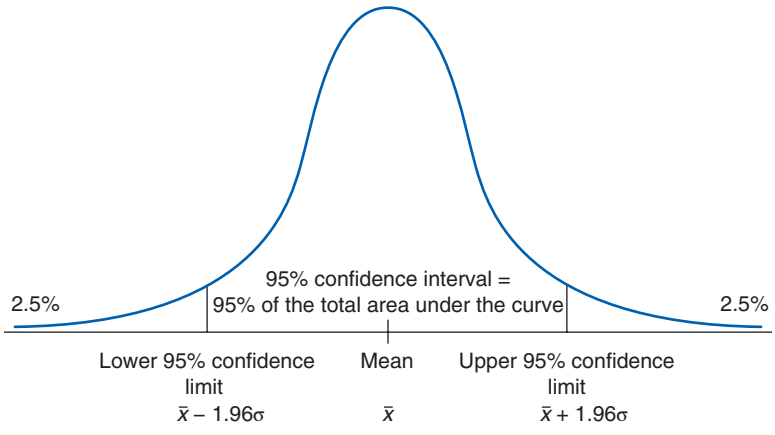
The **standard deviation** is simply the square root of the variance. It is the square root of the average squared deviation of each case from the mean:

$$\sigma_2 = \sqrt{\frac{\sum(Y_i - \bar{Y}_j)^2}{N}}$$

When the standard deviation is calculated from sample data, the denominator is supposed to be $N - 1$ rather than N, an adjustment that has no discernible effect when the number of cases is reasonably large. You also should note the use of *squared* deviations in the formula accentuates the impact of relatively large deviations because squaring a large number makes that number count much more.

The standard deviation has mathematical properties that increase its value for statisticians. You already learned about the **normal distribution** in Chapter 5. A normal distribution is a distribution that results from chance variation around a mean. It is symmetric and tapers off in a characteristic shape from its mean. If a variable is normally distributed, 68% of the cases will lie between plus and minus 1 standard deviation from the distribution's mean, and 95% of the cases will lie between plus and minus 1.96 standard deviations from the mean (see Figure 9.10).

FIGURE 9.10 ■ The Normal Distribution



This correspondence of the standard deviation to the normal distribution enables us to infer how confident we can be that the mean (or some other statistic) of a population sampled randomly is within a certain range of the sample mean. This is the logic behind calculating confidence limits around the mean. You learned in Chapter 5 that confidence limits indicate how confident we can be, based on our random sample, that the value of some statistic in the population falls within

a particular range. (The actual value in the population is the *population parameter.*) Now that you know how to compute the standard deviation, it is just a short additional step to computation of the confidence limits around a mean. There are just four more steps:

1. Calculate the standard error. This is the estimated value of the standard deviation of the sampling distribution from which your sample was selected. $SE = \sigma/\sqrt{(n-1)}$ In words, divide the standard deviation of the sample by the square root of the number of cases in the sample minus one.
2. Decide on the degree of confidence you want to have that the population parameter falls within the confidence interval you compute. It is conventional to calculate the 95%, 99%, or even the 99.9% confidence limits around the mean. Most often, the 95% confidence limits are used, so I will show the calculation for this estimate.
3. Multiply the value of the SE × 1.96. This is because 95% of the area under the normal curve falls within ±1.96 standard deviation units of the mean.
4. Add and subtract the number calculated in (3) from the sample mean. The resulting numbers are the upper and lower confidence limits.

If you had conducted these steps for age with the 2020 GSS data, you would now be able to report, "Based on the GSS 2020 sample, I can be 95% confident that the true mean age in the population of adults 18 and older is between 51.78 and 53.38" (52.58 –.406* 1.96 and 52.58 +.406 * 1.96). When you read in media reports about polling results that the "margin of error" was ±3 points (or whatever value), you'll now know the pollster was simply providing 95% confidence limits for the statistic.

CROSS-TABULATING VARIABLES

Most data analyses focus on relationships between variables to test hypotheses or just to describe or explore relationships. For each of these purposes, we must examine the association between two or more variables. **Cross-tabulation (crosstab)** is one of the simplest methods for doing so. A **cross-tabulation**, or **contingency table**, displays the distribution of one variable for each category of another variable; it can also be termed a *bivariate distribution.* You can also display the association between two variables in a graph; we will see an example in this section. In addition, crosstabs provide a simple tool for statistically controlling one or more variables while examining the associations between others. In the next section, you will learn how crosstabs used in this way can help test for spurious relationships and evaluate causal models. We will examine several trivariate tables.

Constructing Contingency Tables

Cross-tabulation is a useful method for examining the relationship between variables that have a limited number of categories. For most analyses, 10 categories is a reasonable upper limit, but even 10 is too many unless you have a pretty large number of cases (more than 100). If you wish to include in a crosstab a variable with many categories or one that varies along a continuum with many values, you should first recode the values of that variable to a smaller number. For example, you might recode the values of a 5-point index to just *high* (*representing scores of 4 and 5*), *medium* (*3*), and *low* (*1 and 2*). You might recode the numerical values of age to 10-year intervals. Figure 9.11 provides examples.

FIGURE 9.11 ■ Examples of Recoding

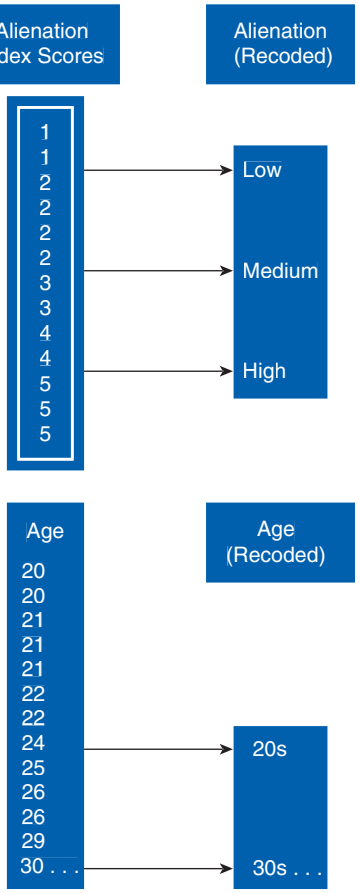

Alienation Index Scores
Alienation (Recoded)
1 1 2 2 2 2 3 3 4 4 5 5 5
Low
Medium
High
Age
Age (Recoded)
20 20 21 21 21 22 22 24 25 26 26 29 30 . . .
20s
30s . . .

Table 9.9 displays the cross-tabulation of likelihood of voting in the 2020 presidential election by family income, so we can test the hypothesis: *likelihood of voting increases with family income*. To simplify the table a bit, I recoded the five categories of likelihood of voting (Table 9.2) into three categories and the values of family income to four categories. The table is presented first with frequencies ("cell counts") and then again with percentages. In both parts of this table, the *body* of the table is the part between the row and column labels and the row and column totals. The cells of the table are defined by combinations of row and column values. Each cell represents cases with a unique combination of values of the two variables, corresponding to that particular row and column. The **column marginal distribution** of the table is underneath (the *column marginals; row marginals* not shown). This is just the frequency distribution for the independent variable (in number of cases, percentages, or both), considered separately. The independent variable is usually the column variable; the dependent variable then is the row variable.

The table in the upper panel of Table 9.9 shows the number of cases with each combination of values of likelihood of voting and family income. In it, you can see that 215 of those earning less than $30,000 per year were extremely likely to vote, whereas 83 were only somewhat likely to vote and 64 were not likely to vote. On the other hand, 288 of those whose family earned $110,000 or more were extremely likely to vote, compared to 34 and 18 of these high-income respondents who were somewhat or not likely to vote, respectively.

TABLE 9.9 ■ Cross-Tabulation of Likelihood of Voting in 2020 by Family Income: Cell Counts and Percentages

	Family Income			
Likelihood of Voting	**< $30,000**	**$30,000–$59,999**	**$60,000–$109,999**	**$110,000+**
	Cell Counts			
Extremely likely	215	282	355	288
Somewhat likely	83	70	52	34
Not likely	64	45	42	18
Total (*n*)	(362)	(397)	(449)	(340)
	Percentages			
Extremely likely	59.4%	71.0%	79.1%	84.7%
Somewhat likely	22.9	17.6	11.6	10.0
Not likely	17.7	11.3	9.4	5.3
Total	100%	100%	100%	100%

Source: General Social Survey, National Opinion Research Center 2020.

It is hard to look at a table in this form (the upper part of Table 9.9), with just the numbers of cases in each cell, and determine whether there is a relationship between the two variables. We need to convert the cell frequencies into **percentages**, as in the table in the lower panel of Table 9.9. This table presents the data as percentages within the categories of the independent variable (the column variable, in this case). In other words, the cell frequencies have been converted into percentages of the column totals (the *n* in each column). For example, in Table 9.9, the number of people in families earning less than $30,000 who were extremely likely to vote is 215 out of 362, or 59.4%. Because the cell frequencies have been converted to percentages of the column totals, the numbers add up to 100 in each column but not across the rows.

To read the percentage table (the bottom panel of Table 9.9), compare the percentage distribution of likelihood of voting across the columns, starting with the lowest income category (in the left column) and moving from left to right. You see that as income increases, the percentage who were "extremely likely" to vote also rises, from 59% (rounding off) of those with annual family incomes less than $30,000 (in the first cell in the first column), to 71% of those with family incomes between $30,000 and $59,999, then 79% of those with family incomes from $60,000 to $109,999, and then up to 85% of those with family incomes of $110,000 or more (the last cell in the first row in the body of the table). At the same time, the percentages in both the "moderately likely" and "not likely" to vote categories decline by about 12 percentage points from the lowest income category ("< $30,000") to the highest income category ("$110,000 or more"). This result is consistent with the hypothesis.

When a table is converted to percentages, usually just the percentage in each cell should be presented, not the number of cases in each cell. Include 100% at the bottom of each column (if the independent variable is the column variable) to indicate that the percentages add up to 100, as well as the *base number* (*n*) for each column (in parentheses). If the percentages add up to 99 or 101 because of rounding error, just indicate this in a footnote.

Follow these rules when you create and then read a percentage table:

1. Make the independent variable the column variable and the dependent variable the row variable.
2. Percentage the table column by column, on the column totals. The percentages should add to 100 (or perhaps 99 or 101, if there has been a slight error due to rounding) in each column.
3. Compare the distributions of the dependent variable (the row variable) across each column.

But the independent variable does not *have* to be the column variable; what is critical is to be consistent within a report or paper. You will find in published articles and research reports some percentage tables in which independent variable and dependent variable positions are reversed. If the independent variable is the row variable, we percentage the table on the row totals (the *n* in each row), and so the percentages total 100 across the rows. Let's examine Table 9.10, which is percentage on the row variable: age. When you read the table in Table 9.10, you find that 52.8% of those in their 20s were extremely likely to vote (while 21.1% were moderately and 26.1% were not likely to vote), compared with 65.7% and 69.0% of those in their 30s and 40s, respectively, 76.9% and 77.8% of those in their 50s and 60s, and 83.7% in their 70s, and then 78.3% of those 80 or older. So likelihood of voting tended to increase with age.

TABLE 9.10 ■ Cross-Tabulation of Likelihood of Voting in 2020 by Age (row percentages)

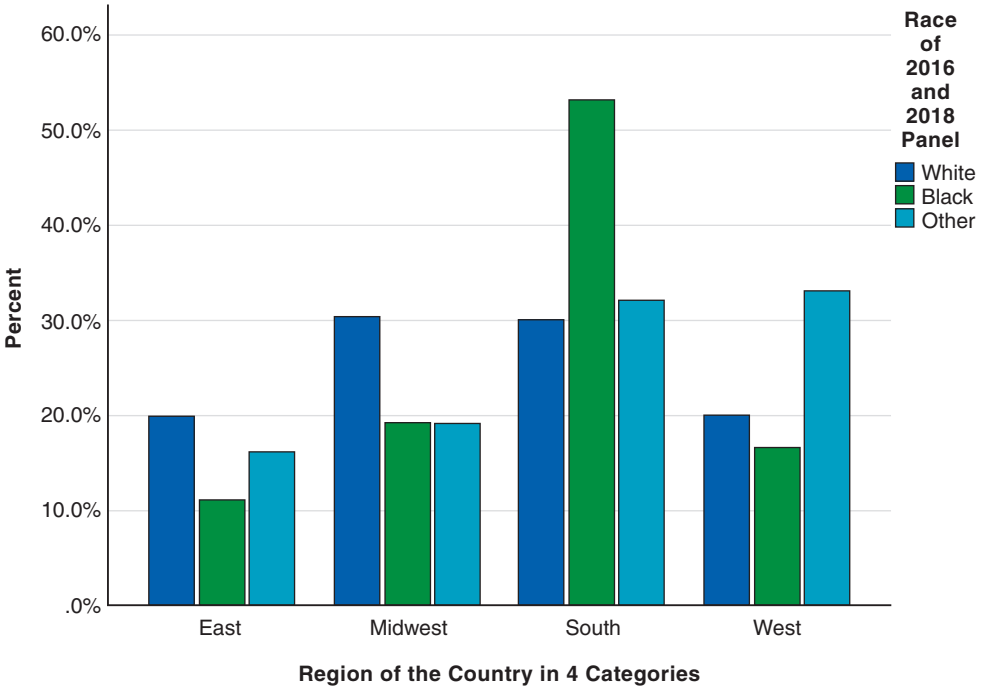
	Likelihood of Voting				
Age	Extremely	Moderately	Not Likely	Total (%)	(*n*)
20–29	52.8%	21.1	26.1	100.0%	(180)
30–39	65.7%	18.4	15.9	100.0%	(283)
40–49	69.0%	18.5	12.5	100.0%	(271)
50–59	76.9%	14.2	8.9	100.0%	(303)
60–69	77.8%	14.3	7.9	100.0%	(329)
70–79	83.7%	12.5	3.8	100.0%	(264)
80 or older	78.3%	14.1	7.6	100.0%	(92)

Source: General Social Survey, National Opinion Research Center 2020.

Graphing Association

Graphs provide an efficient tool for summarizing relationships between variables. Figure 9.12 displays the relationship between race and region in graphic form. It shows that more than half of the Black population in the U.S. who participated in the GSS2020 survey is in the South, whereas almost two thirds of "other" races live in the South and West and 60% of the white population lives in the Midwest and South. The East (New England and the Middle Atlantic states) is the least populous region.

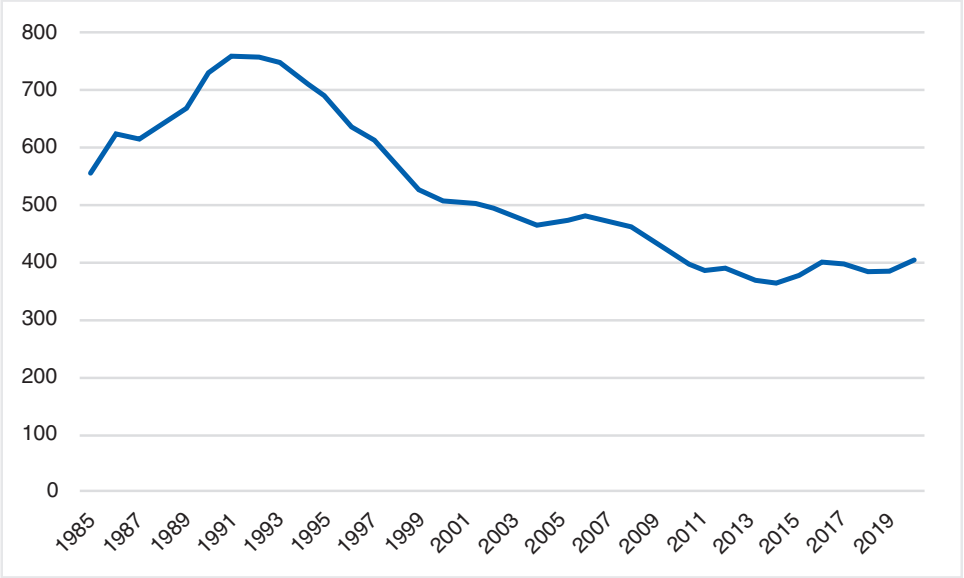
FIGURE 9.12 ■ Race by Region of the United States



Source: General Social Survey, National Opinion Research Center 2020.

Another good example of the use of graphs to show relationships is provided by a graph of violent crimes per 100,000 people from 1985 to 2020 as measured in the *FBI Uniform Crime Reports* (Figure 9.13). The violent crime rate rose from 1985 to 1991, then began a sharp decline that lasted until 2014. After 2014, the violent crime rate rose a bit and remained relatively stable

FIGURE 9.13 ■ Violence in America, 1985–2020



Source: Crime Data Explorer, Federal Bureau of Investigation, 2021. https://crime-data-explorer.app.cloud.gov/pages/explorer/crime/crime-trend.

to 2020 (it increased in 2021). Because the crime rate is measured at the interval–ratio level, the graph can represent the variation over time with a continuous line.

Describing Association

A cross-tabulation table reveals four aspects of the association between two variables:

- *Existence.* Do the percentage distributions vary at all between categories of the independent variable?
- *Strength.* How much do the percentage distributions vary between categories of the independent variable? In most analyses, the analyst would not pay much attention to differences of less than 10 percentage points between categories of the independent variable.
- *Direction.* For quantitative variables, do values on the dependent variable tend to increase or decrease with an increase in value on the independent variable?
- *Pattern.* For quantitative variables, are changes in the percentage distribution of the dependent variable regular (simply increasing or decreasing), or do they vary (perhaps increasing, then decreasing, or perhaps gradually increasing, then rapidly increasing)?

Looking back at Table 9.9, an association exists; it is moderately strong (the difference in percentages between those who were extremely likely to vote in the first and last column is 25.3 percentage points); and the direction of association between likelihood of voting and family income is positive. The pattern in this table is termed **monotonic**. In a monotonic relationship, the value of cases consistently increases (or decreases) on one variable as the value of cases increases on the other variable. This is also largely true in the association between age and likelihood of voting in Table 9.10, although the pattern reverses in the highest age category. By contrast, the relationship charted in Figure 9.12 between the violent crime rate and year is **curvilinear**. The rate increases over the first five years displayed and then decreases, until the increase in 2021.

The relationship between the measure of interpersonal trust and voting appears in Table 9.11. There is an association in the direction I hypothesized: 81.4% of those who believe that people can be trusted were extremely likely to vote, compared with 66.5% of those who believe that people cannot be trusted. Because the independent variable is dichotomous, there is no

TABLE 9.11 ■ Likelihood of Voting in 2020 by Interpersonal Trust*

Can People Be Trusted?		
Likelihood of Voting	**Yes**	**No**
Extremely likely	81.4%	66.5%
Moderately likely	12.2	19.2
Not likely	6.4	14.3
Total	100.0%	100.0%
(*n*)	(485)	(691)

*Those responding "it depends" (1.2% [*n* = 15] of the sample) were recoded to "Yes."

Source: General Social Survey, National Opinion Research Center 2020.

more pattern to the association. (Comparing the column percentages in the second and third rows gives the complementary picture of less likelihood of voting for those who do not trust others, compared to those who do.)

Table 9.12, by contrast, gives no evidence of an association between gender and likelihood of voting in 2020. The difference between the percentage of men and women who were extremely likely to vote is only 2.5 percentage points.

TABLE 9.12 ■ Likelihood of Voting in 2020 by Gender

Gender		
Voting	**Male**	**Female**
Extremely likely	71.2%	73.7%
Moderately likely	17.1	15.3
Not likely	11.8	11.1
Total	100.0%	100.0%
(*n*)	(739)	(983)

Source: General Social Survey, National Opinion Research Center 2020.

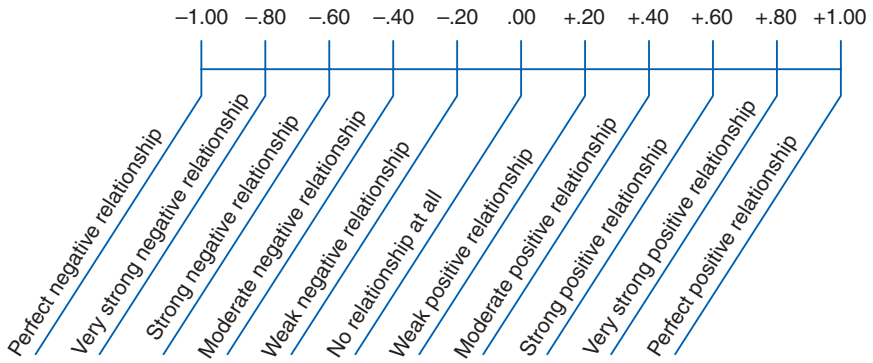
Evaluating Association

You will find when you read research reports and journal articles that social scientists usually make decisions about the existence and strength of association based on more statistics than just a cross-tabulation table.

A **measure of association** is a type of descriptive statistic used to summarize the strength of an association. There are many measures of association, some of which are appropriate for variables measured at particular levels. One popular measure of association in cross-tabular analyses with variables measured at the ordinal level is **gamma**. As with many measures of association, the possible values of gamma vary from -1, meaning the variables are perfectly associated in an inverse direction; to 0, meaning there is no association of the type that gamma measures; to +1, meaning there is a perfect positive association of the type that gamma measures.

Figure 9.14 provides a rough guide to interpreting the value of a measure of association like gamma that can range from 0 to -1 and +1. For example, if the value of gamma is -.23, we could

FIGURE 9.14 ■ A Guide to Interpreting Strong and Weak Relationships



Source: Frankfort-Nachmias and Leon-Guerrero (2006:230). Reprinted with permission from SAGE Publications, Inc.

say there is a *weak negative relationship* between the two variables. If the value of gamma is +.61, we could say there is a *strong positive relationship* between the two variables. A value of 0 always means there is no relationship (although this really means there is no relationship that this statistic can identify). This "rough guide" to interpretation must be modified for some measures of association, and your interpretations must consider the results of previous research and the methods you used to collect your data. For now, however, this rough guide will get you further along in your statistical interpretations.

Inferential statistics are used in deciding whether it is likely that an association exists in the larger population from which the sample was drawn. Even when the association between two variables is consistent with the researcher's hypothesis, it is possible that the association was just caused by the vagaries of sampling on a random basis (of course, the problem is even worse if the sample is not random). It is conventional in statistics to avoid concluding that an association exists in the population from which the sample was drawn unless the probability that the association was due to chance is less than 5%. In other words, a statistician typically will not conclude that an association exists between two variables unless they can be at least 95% confident the association was not due to chance. This is the same type of logic you learned about earlier in this chapter, which introduced the concept of 95% confidence limits for the mean. Estimation of the probability that an association is not due to chance will be based on one of several inferential statistics, **chi-square** being the one used in most cross-tabular analyses. The probability is customarily reported in a summary form such as $p < .05$, which can be translated as "The probability that the association was due to chance is less than 5 out of 100 (5%)." The chi-square test for all the cross-tabulation tables presented so far in this chapter—except Table 9.13—indicated the probability that the association was due to chance is less than.1 out of 100 (1/10th of 1%, or $p < .01$). The association in Table 9.13 was not statistically significant (in fact, the odds that the association between sex and likelihood of voting *was* due to chance were about 50/50).

Tables 9.13 and 9.14 will help you understand the meaning of chi-square, without getting into the details of how it is calculated. Let's propose, as our "null hypothesis," that trust in other people has no association with family income. In that case, trust in people would be the same percentage for all columns in the table—the same as for the overall sample across all four income categories (41.8%). So if there were no association, we would expect based on chance that 41.8% of the 265 people with family income below $30,000 (the first column) will say they can trust other people. Because 41.8% of 265 equals 110, that is the number of people we would "expect" to be trusting and of low income if only chance factors were at work. This is the *expected count,* and it differs from the *actual count* of 72, leaving a *residual* of -38.8. This process is repeated with respect to each cell of the table. (If the dependent variable is a dichotomy, as in Table 9.13, the residuals in the second category of the dichotomy will necessarily be the negative of the residuals in the first category of the dichotomy.)

You can see in Table 9.14 how much the actual percentages of respondents in each income category differ from the marginal percentage of 41.8. The calculation of chi-square reflects this variation, as the larger the deviations of the expected from the observed counts in the various table cells, the less likely it is that the association is due only to chance. Chi-square is calculated with a formula that combines the residuals in each cell. SPSS then compares the value of chi-square to a reference table that indicates how likely it is in a table of the given size that this value could have been obtained based on chance, given the "degrees of freedom" (*df*) in the table [(the number of rows -1) × (the number of columns -1)]. In the crosstab of family income and trust, the value of chi-square was 61.1, and the probability that a chi-square value of this

magnitude was obtained based on chance was less than 1 in 1,000 ($p < .001$). We could therefore be quite confident that an association between these two variables exists in the entire U.S. adult population that the sample was selected from.

TABLE 9.13 ■ Determining the Value of Chi-Square (actual - expected counts = residuals)

	Family Income			
People Can Be Trusted	<$30,000	$30,000–$59,999	$60,000–$109,999	$110,000+
Yes	72 - 110.8 = -38.8	98 - 114.6 = -16.6	144 - 132.1 = 11.9	146 - 102.5 = 43.5
No	193 -1 54.2 = 38.8	176 - 159.4 = 16.6	172 - 183.9 = -11.9	99 - 142.5 = -43.5

$X2 = 61.1$, $df = 3$, $p < .001$.

Source: Based on output from General Social Survey, National Opinion Research Center 2020.

TABLE 9.14 ■ Cross-Tabulation of Interpersonal Trust by Income

	Family Income				
People Can Be Trusted	< $30,000	$30,000–$59,999	$60,000–$109,999	$110,000+	Total
Yes	27.2%	35.8%	45.6%	59.6%	41.8%
No	72.8	64.2	54.4	40.4	58.2
Total	100%	100%	100%	100%	100%
(*n*)	(265)	(274)	(316)	(245)	(1100)

Source: General Social Survey, National Opinion Research Center 2020.

When the analyst feels reasonably confident (at least 95% confident) that an association was not due to chance, it is said the association is statistically significant. **Statistical significance** means an association is not likely to result from chance, according to some criterion set by the analyst. Convention (and the desire to avoid concluding that an association exists in the population when it doesn't) dictates that the criterion be a probability less than 5%.

But statistical significance is not everything. You may remember from Chapter 5 that sampling error decreases as sample size increases. For this same reason, an association is less likely to appear based on chance in a larger sample than in a smaller sample. In a table with more than 1,000 cases, such as those involving the 2020 GSS panel sample, the odds of a chance association are often very low indeed. Even weak associations can be statistically significant with such a large random sample, which means the analyst must be careful not to assume that just because a statistically significant association exists, it is therefore important. In a large sample, an association may be *statistically* significant but still be too weak to be *substantively* significant. All this boils down to another reason for evaluating carefully the strength as well as the existence of an association.

Controlling for a Third Variable

Cross-tabulation can also be used to study the relationship between two variables while controlling for other variables. We will focus our attention on controlling for a third variable in this section, but I will say a bit about controlling for more variables at the section's end. We will examine three different uses for three-variable cross-tabulation: (1) identifying an intervening variable, (2) testing a relationship for spuriousness, and (3) specifying the conditions for a relationship. Each type of three-variable crosstab helps strengthen our understanding of the "focal relationship" involving our dependent and independent variables (Aneshensel 2002). Testing a relationship for possible spuriousness helps meet the nonspuriousness criterion for causality. Identifying an intervening variable can help chart the causal mechanism by which variation in the independent variable influences variation in the dependent variable. Specifying the conditions when a relationship occurs can help improve our understanding of the nature of that relationship and suggest future research that can focus on the effect of these conditions.

All three uses for three-variable cross-tabulation are aspects of **elaboration analysis**: the process of introducing control variables into a bivariate relationship to better understand the relationship (Davis 1985; Rosenberg 1968). We will examine the gamma and chi-square statistics for each table in this analysis.

Intervening Variables

We will first complete our test of one of the implications of the causal model of voting in Figure 9.1: that interpersonal trust intervenes in the relationship between social status and likelihood of voting. You already have seen that both income (one of our social status indicators) and trust in people are associated with the likelihood of voting. Both relationships are predicted by the model: so far, so good. You can also see in Table 9.14 that trust is related to income: Higher income is associated positively with the belief that people can be trusted ($p < .001$). Another prediction of the model is confirmed. But to determine whether the trust variable is an intervening variable in this relationship, we must determine whether it explains (transmits) the influence of income on trust. We therefore examine the relationship between income and likelihood of voting while controlling for the respondent's belief that people can be trusted.

According to the causal model (see Figure 9.1), income (social status) influences likelihood of voting (political participation) by influencing trust in people (our substitute for efficacy), which, in turn, influences likelihood of voting. We can evaluate this possibility by reading the two subtables in Table 9.15. **Subtables** like this describe the relationship between two variables within the discrete categories of one or more other control variables. The control variable in Table 9.16 is trust in people, and the first subtable is the income-likelihood of voting crosstab for only those respondents who believe people can be trusted. The second subtable is for those respondents who believe people can't be trusted. They are called subtables because together they make up the table in Table 9.14. If trust in ordinary people intervened in the income–likelihood of voting relationship, then the effect of controlling for this third variable would be to eliminate or at least substantially reduce this relationship—the distribution of likelihood of voting would be the same for every income category in both subtables in Table 9.15.

A quick inspection of the subtables in Table 9.15 reveals trust in people does not intervene in the relationship between family income and likelihood of voting. In both subtables, the likelihood of respondents planning to vote rose with their incomes. Of course, this finding does not necessarily mean the causal model was wrong. This one measure is a measure of trust in people,

TABLE 9.15 ■ Likelihood of Voting in 2020 by Family Income by Interpersonal Trust

	Family Income			
People Can Be Trusted				
Likelihood of voting	**< $30,000**	**$30,000–$59,999**	**$60,000–$109,999**	**$110,000+**
Extremely likely	66.7%	79.6%	83.2%	89.4%
Moderately likely	23.2	12.9	10.5	7.0
Not likely	10.1	7.5	6.3	3.5
	100%	100%	100%	100%
(*n*)	(69)	(93)	(143)	(142)
People Cannot Be Trusted				
Extremely likely	58.0%	69.6%	70.5%	73.5%
Moderately likely	24.1	19.9	13.9	15.3
Not likely	17.8	10.5	15.7	11.2
Total	100%	100%	100%	100%
(*n*)	(174)	(171)	(166)	(98)

Source: General Social Survey National Opinion Research Center 2020.

which is not the same as the widely studied concept of political efficacy; a better measure, from a different survey, might function as an intervening variable. But for now, we should be less confident in the model. We can also see in this table that those who believe people can be trusted are more likely to vote than those who believe people cannot be trusted at every income level. So variation in interpersonal trust does help to explain likelihood of voting, but it does so in addition to the influence of family income on likelihood of voting, not as part of the process by which family income influences likelihood of voting.

Extraneous Variables

Another reason for introducing a third variable into a bivariate relationship is to see whether that relationship is spurious because of the influence of an extraneous variable (see Chapter 6)—a variable that influences both the independent and dependent variables, creating an association between them that disappears when the extraneous variable is controlled. Ruling out possible extraneous variables will help strengthen considerably the conclusion that the relationship between the independent and dependent variables is causal, particularly if all the variables that seem likely to have the potential for creating a spurious relationship can be controlled.

One variable that might create a spurious relationship between family income and voting is education. You have already seen that the likelihood of voting increases with family income. Is it not possible, though, that this association is spurious because of the effect of education? Education, after all, is associated with both family income and voting, and we might surmise that it is what students learn in school about civic responsibility that increases voting, not family

income itself. Figure 9.15 diagrams this possibility, and Table 9.16 shows the bivariate associations between education and likelihood of voting and education and income. As the model in Figure 9.15 predicts (the diagonal arrows), education is associated with both income and voting. So far, so good.

If education creates a spurious relationship between income and voting, there should be no association between income and voting after controlling for education. Because we are using crosstabs, this means there should be no association in any of the income–voting subtables for any value of education.

FIGURE 9.15 ■ A Causal Model of a Spurious Effect

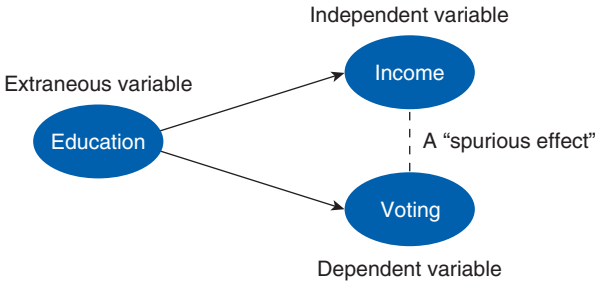


TABLE 9.16 ■ Likelihood of Voting in 2020 by Education and Income by Education

	Education			
Likelihood of Voting	**High School or Less**	**Some College**	**College Graduate**	**Graduate Work**
Extremely	58.8%	73.1%	80.6%	85.1%
Moderately	21.8	16.5	12.2	10.7
Not likely	19.4	10.4	7.2	4.2
Total	100%	100%	100%	100%
(*n*)	(546)	(442)	(377)	(355)
	Education			
Family Income	**High School Graduate**	**Some College**	**College Graduate**	**Graduate Work**
< $30,000	40.4%	29.1%	11.0%	8.1%
$30,000– $59,999	28.1	29.1	20.4	24.6
$60,000– $109,999	24.1	29.1	32.0	30.3
$110,000+	7.4	12.8	36.6	37.0
Total	100%	100%	100%	100%
(*n*)	(498)	(413)	(363)	(346)

Source: General Social Survey, National Opinion Research Center 2020.

The trivariate cross-tabulation in Table 9.17 shows the relationship between likelihood of voting and income (see Table 9.9) is not spurious because of the effect of education; if it were, an association between voting and family income wouldn't appear in any of the subtables. However, the strength of the association between likelihood of voting and family income is somewhat weaker in the subtables (average subtable gamma = -.224, compared to the complete table [9.9] gamma = -.32), so we can conclude that some of the apparent effect of family income on likelihood of voting is due to its association with education.

TABLE 9.17 ■ Voting in 2020 by Income and Education

	Family Income			
Likelihood of Voting	**< $30,000**	**$30,000–$59,999**	**$60,000–$109,999**	**$110,000+**
Education = High School or Less				
Extremely	53.8%	55.4%	71.6%	75.7%
Moderately	25.3	22.3	13.8	13.5
Not likely	21.0	22.3	14.7	10.8
Total	100%	100%	100%	100%
(*n*)	(186)	(130)	(116)	(37)
Education = Some College				
Extremely	60.5%	79.3%	80.2%	78.8%
Moderately	22.8	15.5	13.8	5.8
Not likely	16.7	5.2	6.0	15.4
Total	100%	100%	100%	100%
(*n*)	(114)	(116)	(116)	(52)
Education = College Graduate				
Extremely	72.2%	79.2%	77.9%	85.9%
Moderately	16.7	13.9	10.6	10.9
Not likely	11.1	6.9	11.5	3.1
Total	100%	100%	100%	100%
(*n*)	(36)	(72)	(113)	(128)
Education = Graduate Work				
Extremely	76.9%	77.2%	87.5%	88.6%
Moderately	15.4	16.5	7.7	9.8
Not likely	7.7	6.3	4.8	1.6
Total	100%	100%	100%	100%
(*n*)	(26)	(79)	(104)	(123)

Source: General Social Survey, National Opinion Research Center 2020.

Specification

By adding a third variable to an evaluation of a bivariate relationship, the data analyst can also specify the conditions under which the bivariate relationship occurs. A **specification** occurs when the association between the independent and dependent variables varies across the categories of one or more other control variables. If you look back at Table 9.15, you can see there is a modest difference in the strength of the family income and likelihood of voting association in the subtables (as reflected in the value of gamma, which is -.327, $p < .001$ in the first subtable and -.155, $p < .01$ in the second). This indicates that family income has more of an effect on likelihood of voting among respondents who believe people can be trusted than among those who do not believe people can be trusted.

Tables 9.18 and 9.19 reveal a more pronounced specification pattern. Approval of Donald Trump as president dropped as education rose among the 2020 GSS respondents overall, from 43.8% among those with no more than a high school education to 28.1% among those with some graduate education after college (see Table 9.18).

TABLE 9.18 ■ Approval of Trump as President in 2020 by Education

Approval of Trump by Education				
Education				
Approval of Trump	High School or Less	Some College	College Graduate	Graduate Work
Approve	43.8%	37.3%	30.8%	28.1%
Disapprove	56.2	62.7	69.2	71.9
Total	100%	100%	100%	100%
(*n*)	(557)	(445)	(380)	(363)

Gamma =.20, $p < .001$

However, educational level only made a difference among white Americans, so race specified the effect of education on approval of Trump. Approval of Trump decreased as education increased for whites, from a high of 54.8% among those with no more than a high school education to a low of 30.7% among those with at least some graduate work (see Table 9.19). Among Black Americans, approval of Trump varied only slightly by education, from 7.4% among those with no more than a high school education to 5.1% of those with at least some graduate work. Among those identifying as of some other race, more than one quarter approved of Trump, but this approval did not vary by educational level.

I should add one important caution about constructing tables involving three or more variables. Because the total number of cells in the subtables becomes large as the number of categories of the control (third) variable increases, the number of cases that each cell percentage is based on will become correspondingly small. This effect has two important consequences. First, the number of comparisons that must be made to identify the patterns in the table as a whole becomes substantial—and the patterns may become too complex to make much sense of them. Second, as the number of cases per category decreases, the odds that the distributions within

TABLE 9.19 ■ Approval of Trump as President in 2020 by Education and Race

Approval of Trump by Education by Race				
Approval of Trump	**High School or Less**	**Some College**	**College Graduate**	**Graduate Work**
Race = White				
Approve	54.8%	45.3%	35.3%	30.7%
Disapprove	45.2	54.7	64.7	69.3
Total	100%	100%	100%	100%
(*n*)	(407)	(329)	(289)	(303)
Gamma =.29, p < .001				
Race = Black				
Approve	7.4%	7.2%	6.1%	5.1%
Disapprove	92.6	92.8	93.9	94.9
Total	100%	100%	100%	100%
(*n*)	(95)	(83)	(49)	(37)
Gamma =.09, NS (p > .05)				
Race = Other				
Approve	25.5%	33.3%	28.6%	33.3%
Disapprove	74.5	66.7	71.4	66.7
Total	100%	100%	100%	100%
(*n*)	(244)	(166)	(117)	(102)
Gamma = -.08, NS (p > .05)				

each category vary because of chance become greater. This problem of having too many cells and too few cases can be lessened by making sure the control variable has only a few categories and by drawing a large sample, but often neither of these steps will be sufficient to resolve the problem completely.

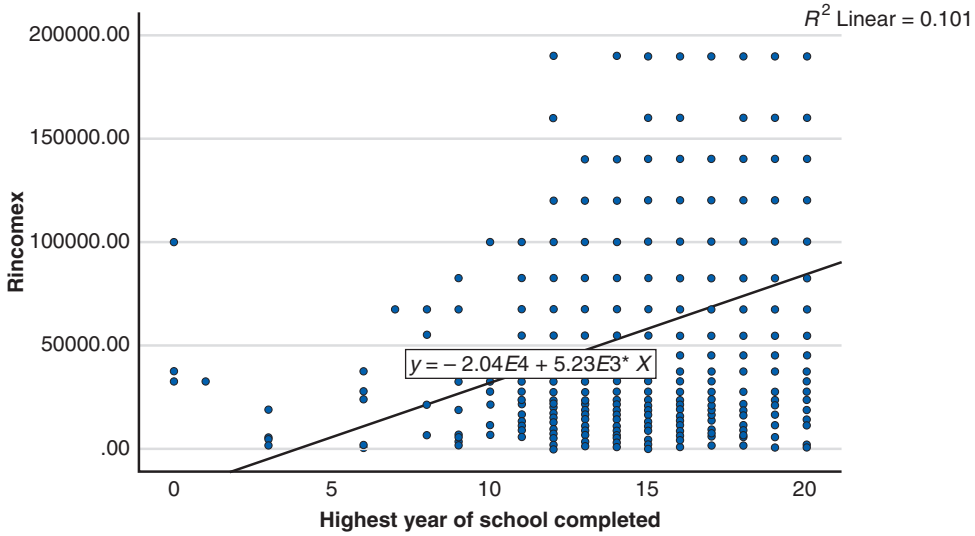
REGRESSION ANALYSIS

My goal in introducing you to cross-tabulation has been to help you think about the association between variables and to give you a relatively easy tool for describing association. To read most statistical reports and to conduct more sophisticated analyses of social data, you will have to extend your statistical knowledge. Many statistical reports and articles published in social science journals use a statistical technique called **regression analysis** or **correlational analysis** to describe the association between two or more quantitative variables. The terms actually refer to different aspects of the same technique. Statistics based on regression and correlation are used

frequently in social science, and have many advantages over cross-tabulation—as well as some disadvantages.

I give you only an overview of this approach here. Look at Figure 9.16. It's a plot, termed a *scatterplot*, of the relationship in the GSS 2020 sample between years of education and respondent's income. You can see that I didn't collapse the values of either of these variables into categories, as I had to do to use them in the preceding cross-tabular analysis. Instead, the scatterplot shows the location of each case in the data relative to years of education (the horizontal axis) and respondent's income level (the vertical axis). (Due to the large number of cases and the limited number of values, each point in the scatterplot may represent many cases.)

FIGURE 9.16 ■ Respondent Income by Highest Year of School Completed



Source: General Social Survey, National Opinion Research Center 2020.

You can see the data points in the scatterplot tend to run from the lower left to the upper right of the chart, indicating a positive relationship: The more the years of education, the higher the respondent's income. The line drawn through the points is the regression line. The regression line summarizes this positive relationship between years of education, which is the independent variable (often simply termed *X* in regression analysis), and respondent's income, the dependent variable (often simply termed *Y* in regression analysis). This regression line is the "best fitting" straight line for this relationship—it is the line that lies closest to all the points in the chart, according to certain criteria. But you can easily see that quite a few points are pretty far from the regression line.

How well does the regression line fit the points? In other words, how close does the regression line come to the points? (Actually, it's the square of the vertical distance, on the *y*-axis, between the points and the regression line that is used as the criterion.) The **correlation coefficient**, also called *Pearson's r*, or just *r*, gives one answer to that question. The value of *r* for this relationship is .10, which indicates a weak positive linear relationship (if it were a negative relationship, *r* would have a negative sign). The value of *r* is 0 when there is absolutely no linear relationship between the two variables, and it is 1 when all the points representing all the cases lie exactly on the regression line (which would mean the regression line describes the relationship perfectly).

So the correlation coefficient does for a scatterplot such as this what gamma does for a crosstabulation table: It is a summary statistic that tells us about the strength of the association between the two variables. Values of *r* close to 0 indicate the relationship is weak; values of *r* close to ±1 indicate the relationship is strong—in between there is a lot of room for judgment. You will learn in a statistics course that r^2 is often used instead of *r*.

You can also use correlation coefficients and regression analysis to study simultaneously the association between three or more variables. In such a *multiple regression analysis*, you could test to see whether several other variables in addition to education are associated simultaneously with respondent's income—that is, whether the variables have independent effects on respondent's income. As an example, Table 9.20 presents the key statistics obtained in a multiple regression analysis I conducted with the GSS 2020 data to do just that: I regressed respondent's income (in approximate dollars) on years of schooling, age, sex, and race (dichotomized).

TABLE 9.20 ■ Multiple Regression of Determinants of Respondent's Income

Variable	Beta Coefficient (ß)	Significance Level
Age	.08	$p < .01$
Education	.32	$p < .001$
Sex (female)	–.24	$p < .001$
Race (minority)	–.05	*NS*
R^2	.18	
N	1,077	

Source: General Social Survey, National Opinion Research Center 2020.

First, look at the numbers under the "Beta Coefficient" heading. Beta coefficients are standardized statistics that indicate how strong the linear association is between the dependent variable (respondent's income, in this case) and each independent variable, while the other independent variables are controlled. Like the correlation coefficient (*r*), values of beta range from 0, when there is no linear association, to ±1.0, when the association falls exactly on a straight line. You can see in the beta column that education has a moderate positive independent association with respondent's income, whereas sex (being female) has a negative independent association with respondent's income. Age has a weak positive independent association with respondent's income. The independent association of race with respondent's income (after the other variables are controlled) is not statistically significant. Both sex and race are coded as dichotomies for this analysis, with "female" = 1 and "male" = 0 for sex, and "minority" = 1 and "white" = 0 for race. Thus, the negative coefficient for sex means the higher value of sex (being female rather than male) is associated with lower income. In the "Sig." column, you can see that two of the effects are statistically significant at the .001 level, while age is significant at the .01 level. You learn from the summary statistic, R^2 ("*r*-squared"), that the four independent variables together explain, or account for, 18% of the total variation in respondent's income.

You will need to learn more about when correlation coefficients and regression analysis are appropriate (e.g., both variables must be quantitative, and the relationship has to be linear [not curvilinear, unless some mathematical adjustments are made]). But that's for another time and

place. To learn more about correlation coefficients and regression analysis, you should take an entire statistics course. For now, this short introduction will enable you to make sense of more of the statistical analyses you find in research articles. You can also learn more about these techniques with the tutorials on the text's study site.

ANALYZING DATA ETHICALLY

Using statistics ethically means first and foremost being honest and open. Findings should be reported honestly, and the researcher should be open about the thinking that guided their decision to use particular statistics. It is possible to distort social reality with statistics, and it is unethical to do so knowingly, even when the error is due more to carelessness than deceptive intent.

Summary statistics can easily be used unethically, knowingly or not. When we summarize a distribution in a single number, even in two numbers, we are losing much information. Neither central tendency nor variation describes a distribution's overall shape. And taken separately, neither measure tells us about the other characteristic of the distribution (central tendency or variation). Reports using measures of central tendency should normally also include measures of variation. And we should inspect the shape of any distribution for which we report summary statistics to ensure the summary statistic does not mislead us (or anyone else) because of an unusual degree of skewness.

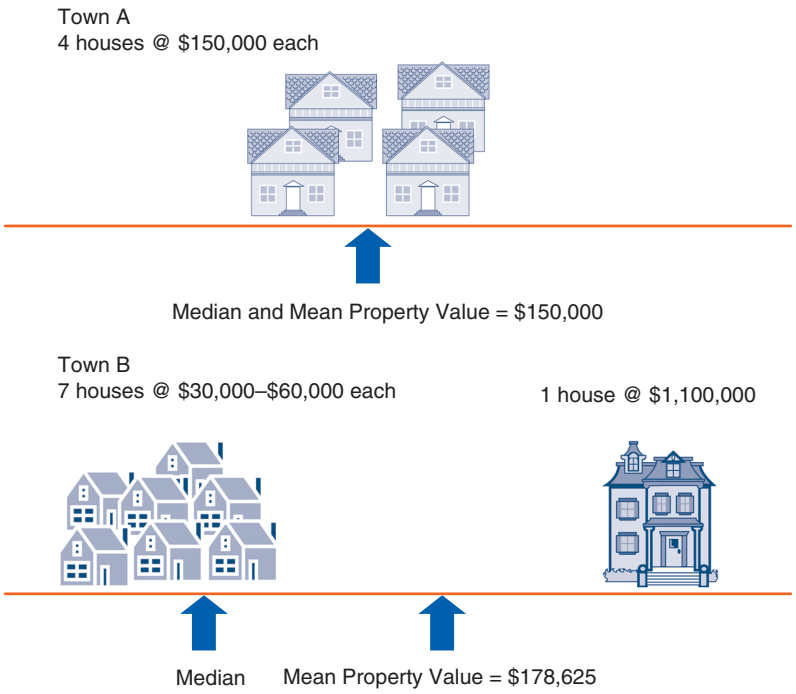
It is possible to mislead those who read statistical reports by choosing summary statistics that accentuate a particular feature of a distribution. You probably realized this potential when I discussed Figure 9.9, so here I've only reproduced part of it. Imagine an unscrupulous realtor trying to convince a prospective home buyer in Town B that it is a community with very high property values, when it actually has a positively skewed distribution of property values (see Figure 9.17). The realtor compares the mean price of homes in Town B to that for Town A (one with a homogeneous mid-priced set of homes) and therefore makes Town B look much better. In truth, the higher mean in Community B reflects a very skewed, lopsided distribution of property values—most residents own small, cheap homes, but one owns a mansion and so the mean is pulled up in the direction of this positive skew. A median would provide a better basis for comparison.

You have already seen that it is possible to distort the shape of a distribution by ignoring some of the guidelines for constructing graphs and frequency distributions. Whenever you need to group data in a frequency distribution or graph, you can reduce the potential for problems by inspecting the ungrouped distributions and then using a grouping procedure that does not distort the distribution's basic shape. When you create graphs, be sure to consider how the axes you choose may change the distribution's apparent shape.

When the data analyst begins to examine relationships between variables in some real data, social science research becomes most exciting. The moment of truth, it would seem, has arrived. Either the hypotheses are supported or they are not. But this is a time to proceed with caution and to evaluate the analyses of others with even more caution. Once large data sets are entered into a computer, it becomes very easy to check out a great many relationships; when relationships are examined between three or more variables at a time, the possibilities become almost endless. In fact, regression analysis (extended to *multiple regression analysis*) allows a researcher to test easily for many relationships at the same time.

This range of possibilities presents a great hazard for data analysis. It becomes tempting to search around in the data until something interesting emerges. Rejected hypotheses are forgotten in favor of highlighting what's going on in the data. It's not wrong to examine data for unanticipated relationships; the problem is that inevitably some relationships between variables will

FIGURE 9.17 ■ Using the Mean to Create a More Favorable Impression


Town A
4 houses @ $150,000 each
Median and Mean Property Value = $150,000
Town B
7 houses @ $30,000–$60,000 each
1 house @ $1,100,000
Median
Mean Property Value = $178,625
"Yes, our lovely town B has a higher average property value than town A."

appear just on the basis of chance association alone. If you search hard and long enough, it will be possible to come up with something that really means nothing.

A reasonable balance must be struck between deductive data analysis to test hypotheses and inductive analysis to explore patterns in a data set. Hypotheses formulated in advance of data collection must be tested as they were originally stated; any further analyses of these hypotheses that involve a more exploratory strategy must be labeled in research reports as such. Serendipitous findings do not need to be ignored, but they must be reported as such. Subsequent researchers can try to test deductively the ideas generated by our explorations.

We also should be honest about the limitations of using survey data to test causal hypotheses. The usual practice for those who seek to test a causal hypothesis with nonexperimental survey data is to test for the association between the independent and dependent variables, controlling for other variables that might possibly create a spurious association. This is what we did by examining the relationship between income and voting while controlling for education (Table 9.17). This three-variable analysis showed that the association between income and voting was not spurious due to the effect of education.

But finding that a hypothesized relationship is not altered by controlling for just one variable does not establish that the association is causal—nor does controlling for two, three, or many more variables. There is always a possibility that some other variable that we did not think to control or that was not even measured in the survey has produced a spurious association between the independent and dependent variables in our hypothesis (Lieberson 1985). We have to think about the possibilities and be cautious in our causal conclusions.

It is also important to understand the statistical techniques we are using and to use them appropriately. In particular, the analyst who uses regression analysis has to make a number of

assumptions about the variables in the analysis; when these assumptions are violated, regression results can be very misleading. (You just might want to rush right out and buy this statistics text at this point: *Social Statistics for a Diverse Society*, 9th edition, by Chava Frankfort-Nachmias, Anna Leon-Guerrero, and Georgiann Davis.

CONCLUSIONS

This chapter has demonstrated how a researcher can describe social phenomena, identify relationships between them, explore the reasons for these relationships, and test hypotheses about them. Statistics provide a remarkably useful tool for developing our understanding of the social world, a tool that we can use both to test our ideas and to generate new ones.

Unfortunately, to the uninitiated, the use of statistics can seem to end debates right there—you can't argue with the numbers. But you now know better than that. The numbers will be worthless if the methods used to generate the data are not valid; and the numbers will be misleading if they are not used appropriately, considering the type of data to which they are applied. And even assuming valid methods and proper use of statistics, there's one more critical step because the numbers do not speak for themselves. Ultimately, it is how we interpret and report the statistics that determine their usefulness.

KEY TERMS

Bar chart (p. 308)
Base number (*N*)
Bimodal (p. 317)
Central tendency (p. 307)
Chi-square (p. 331)
Coding (p. 306)
Combined frequency display (p. 314)
Compressed frequency display
Contingency table (p. 324)
Correlational analysis (p. 338)
Correlation coefficient (p. 339)
Cross-tabulation (crosstab) (p. 324)
Curvilinear (p. 329)
Data cleaning (p. 306)
Data entry (p. 306)
Descriptive statistics (p. 303)
Elaboration analysis (p. 333)
Frequency distribution (p. 311)
Frequency polygon (p. 309)
Gamma (p. 330)
Histogram
Interquartile range (p. 322)
Marginal distribution (p. 325)
Mean (p. 317)
Measure of association (p. 330)
Median (p. 317)
Mode (p. 316)
Monotonic (p. 329)
Normal distribution (p. 323)
Outlier (p. 321)
Percentages (p. 326)
Probability average (p. 316)
Quartiles (p. 322)
Range (p. 321)
Regression analysis (p. 338)
Skewness (p. 307)
Specification (p. 337)
Standard deviation (p. 323)
Statistical significance (p. 332)
Subtables (p. 333)
Unimodal (p. 317)
Variability (p. 307)
Variance (p. 322)

HIGHLIGHTS

LO 9.1 Explain the use of descriptive statistics.

- Descriptive statistics are used to describe the distribution of and relationship between variables.
- Some descriptive statistics describe the results of measuring single variables and help us to construct and evaluate multi-item scales.
- Other descriptive statistics help us describe the association between variables and control for or otherwise account for other variables.
- Descriptive statistics include frequency distributions, graphs, measures of central tendency and variation, and reliability tests.

LO 9.2 Describe the steps of coding and data cleaning that are required to prepare for data analysis.

- Coding is the process of assigning a unique numerical code to each response to survey questions. Coding of all responses should be done before data entry.
  - If the survey is distributed through the Internet or completed directly on a computer, database software can be programmed to recognize the response codes and record them as numbers when the survey responses are entered.
  - However, when responses are recorded on paper-and-pencil questionnaires, they must be entered into a database as a separate process, either through keyboard entry or through scanning a properly prepared form.
- Data cleaning is the process of checking data for errors after the data have been entered into a computer file.
  - Most data entry programs that enter survey responses into a database can be programmed to check each response before it is saved and ensure that it represents a valid response for that variable.
  - When an invalid entry occurs or no response is entered for a required question the program prompts the respondent to try again. This process reduces sharply the possibility of data entry errors.

LO 9.3 Compare the instances when it is appropriate to use graphs to represent statistics to the instances when it is appropriate to use frequency distributions to represent statistics.

- Graphs include bar charts, histograms, and frequency polygons.
  - Bar chart: A graphic for qualitative variables in which the variable's distribution is displayed with solid bars separated by spaces; this is a good tool for displaying the distribution of variables measured at the nominal level. Bar charts can be used to assess central tendency and variability, but not skewness.
  - Histogram: A graphic for quantitative variables in which the variable's distribution is displayed with adjacent bars; this is a good tool for displaying the shape, such as skewness, of the distribution of a quantitative variable.
  - Frequency polygon: A graphic for quantitative variables in which a continuous line connects data points representing the variable's distribution. This is another good choice for displaying the shape of a quantitative variable measured at the interval or ratio level.

- When the goal of a presentation is to convey a general sense of a variable's distribution, particularly when the presentation is to an audience not trained in statistics, a graph is often the best choice.
- Frequency distributions can include combined frequency displays and compressed frequency displays.
  - Combined frequency display: A table that presents together the distributions for a set of conceptually similar variables having the same response categories; common headings are used for the responses.
  - Compressed frequency display: A table that presents cross-classification data efficiently by eliminating unnecessary percentages, such as the percentage corresponding to the second value of a dichotomous variable.
- A frequency distribution can provide more precise information than a graph about the number and percentage of cases in a variable's categories.
- Whichever type of display (graph or frequency distribution) is used, the primary concern of the data analyst is to display accurately the distribution's shape (its central tendency, variability, and skewness [for quantitive variables]) and thus, to show how cases are distributed across the values of the variable.

LO 9.4 Defend the inclusion of both measures of central tendency and measures of variation in data analysis.

- Central tendency is only one aspect of the shape of a distribution—the most important aspect for many purposes but still just a piece of the total picture. A summary of distributions based only on their central tendency can be incomplete, even misleading.
- Analysts who report summary measures of central tendency usually also report a summary measure of variability and sometimes several measures of central tendency, variability, or both, to understand all aspects of the shape of a distribution.
- Measures of central tendency include mode, median, and mean.
  - To choose an appropriate measure of central tendency, the analyst must consider a variable's level of measurement, the skewness of a quantitative variable's distribution, and the purpose for which the statistic is used. In addition, the analyst's personal experiences and preferences inevitably will play a role.
- Measure of variation include range, interquartile range, variance, and standard deviation.
  - To calculate these measures of variation, the variable must be at the interval or ratio level (but many would argue that, like the mean, they can be used with ordinal-level measures, too).
  - Measures of variation do not tell us about the extent to which a distribution is skewed, which we've seen is very important for interpreting measures of central tendency.

LO 9.5 Describe the ways in which cross-tabular analyses can be used to examine the relationship between variables.

- *Constructing contingency tables:* A contingency table can be used to display the distribution of one variable for each category of another variable; this can also be termed a *bivariate distribution*.

- *Graphing association:* Graphs of cross-tabulations can be used to provide an efficient tool for summarizing relationships between variables.
- *Describing association:* A cross-tabulation table reveals four aspects of the association between two variables: existence, strength, direction, and pattern.
- *Evaluating association:* A measure of association is a type of descriptive statistic used to summarize the strength of an association.
  - One popular measure of association in cross-tabular analyses with variables measured at the ordinal level is gamma.
  - In cross-tabular analyses, chi-square is calculated to conclude whether there is under a 5% likelihood that an association in the population from which the sample was drawn is due to chance. If so, the association is considered statistically significant.
- *Controlling for a third variable:* Cross-tabulation can also be used to study the relationship between two variables while controlling for other variables.
  - Three different uses for three-variable cross-tabulation are (1) identifying an intervening variable, (2) testing a relationship for spuriousness, and (3) specifying the conditions for a relationship.
  - Elaboration analysis is the process of introducing a third variable into an analysis to better understand—to elaborate—the bivariate (two-variable) relationship under consideration. Additional control variables also can be introduced.

LO 9.6 Summarize the meaning of the correlation coefficient in a regression analysis.

- The regression line from a regression analysis is the "best fitting" straight line for this relationship—it is the line that lies closest to all the points in the chart, according to certain criteria.
- The correlation coefficient, also called *Pearson's r*, or just *r, tells you how close the regression line comes to the points*.
- The value of *r* is 0 when there is absolutely no linear relationship between the two variables, and it is 1 (or -1) when all the points representing all the cases lie exactly on the regression line (which would mean that the regression line describes the relationship perfectly).
- So the correlation coefficient does for a scatterplot such as this what gamma does for a crosstabulation table: It is a summary statistic that tells us about the strength of the association between the two variables.

LO 9.7 Describe why it is so easy (knowingly or unknowingly) to present misleading information in summary statistics.

- It is possible to mislead those who read statistical reports by choosing summary statistics that accentuate a particular feature of a distribution.
- It is possible to distort the shape of a distribution by ignoring some of the guidelines for constructing graphs and frequency distributions.
- It becomes tempting to search around in the data until something interesting emerges. It's not wrong to examine data for unanticipated relationships; the problem is that inevitably some relationships between variables will appear just on the basis of chance association alone.

- One may be tempted to assume that if a hypothesized relationship is not altered by controlling for just one variable (or two, three, or many more variables), that the association is causal. However, there is always a possibility that some other variable that we did not think to control or that was not even measured in the survey has produced a spurious association between the independent and dependent variables in our hypothesis.

DISCUSSION QUESTIONS

1. I presented in this chapter several examples of bivariate and trivariate cross-tabulations involving family or respondent income as reported in the 2020 GSS. What additional influences would you recommend examining to explain income? Suggest some additional independent variables for bivariate analyses with income as well as several additional control variables to be used in three-variable crosstabs.

2. When should we control... just to be honest? In the evaluation project that I will describe in Chapter 13, I analyzed with some colleagues the effect on cognitive functioning (thinking skills) of living in group homes rather than individual apartments. I found that living in group homes resulted in gains in cognitive functioning, compared with living in individual apartments for persons who had been homeless and diagnosed with serious mental illness. However, this benefit of group homes occurred only for residents who were not substance abusers; substance abusers did not gain cognitively from living in group (or individual) homes (Caplan et al. 2006). Would it have been all right if we had just reported the bivariate association between housing type and change in cognitive functioning? Should social researchers be expected to investigate alternative explanations for their findings? Should they be expected to check to see if the associations they find occur for different subgroups in their samples?

SPSS EXERCISES

1. Develop a description of the basic social and demographic characteristics of the U.S. population in 2020. Examine each characteristic with three statistical techniques: a graph, a frequency distribution, and a measure of central tendency.
   - **a.** From the menu, select Graphs and then Legacy Dialogs and Bar. Select Simple Define [MARITAL_2—Category Axis].
   - **b.** Select Histogram for each of the variables [EDUC_2, CENSUSHHINCOME, ATTEND_2] from the GSS2020 data set.
   - **c.** Describe the distribution of each variable.
   - **d.** Generate frequency distributions and descriptive statistics for these variables. From the menu, select Analyze/ Descriptive Statistics/Frequencies. In the Frequencies window, select MARITAL_2, EDUC_2, CENSUSHHINCOME, ATTEND_2, moving each to the Variable(s) window. Then choose Statistics and choose the mean and median. Then choose Continue (and then OK).
   - **e.** Which statistics are appropriate to summarize the central tendency and variation of each variable? Do the values of any of these statistics surprise you?

2. Propose a variable that might have created a spurious relationship between income [EDUC4] and candidate preference [TRUMP2020], such as race [RACE2020]. Explain your thinking. Propose a variable that might result in a conditional effect of income on voting so that the relationship between income and voting would vary across the categories of the other variable [such as FEMALE]. Test these propositions with three-variable cross-tabulations. Were any supported? How would you explain your findings?

©iStockphoto.com/Sarah J Lee

10 QUALITATIVE METHODS

Russell K. Schutt and Laura R. Golden

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

10.1 Recall the features of qualitative research that most distinguish it from experimental and survey research designs

10.2 Contrast the styles of qualitative research

10.3 Compile a list of the important considerations when conducting participant observation

10.4 Compare the process of intensive interviewing to the process of interviewing in survey research

10.5 Assess how focus group conversations are similar to and different from ordinary, non-research conversations

10.6 Argue whether it makes sense to be concerned with generalizability in qualitative research

10.7 Identify the major ethical challenges faced by qualitative researchers

On Sunday, August 29, 2021, Hurricane Ida made landfall near Port Fouchon, Louisiana. A Category 4 storm, Ida brought winds up to 150 miles per hour, high storm surges, and extreme rain, knocking out power to over 1 million Louisiana residents—including the entire city of New Orleans (Reckdahl et al. 2021; National Oceanic and Atmospheric Administration [NOAA], 2021a). The remnants of Hurricane Ida then made their way up the East Coast, causing historic flooding in New Jersey and New York. With initial estimates of the costs of the damage above $60 billion, Ida was one of the top five costliest hurricanes on record in the United States since 1980 (NOAA 2021b). One new Jersey resident described the trauma of the Hurricane Ida flooding (Kaysen 2021):

> You think of your home as this safe place where it's yours and you're invested in it. Now that this has happened, knowing how bad it can get in our house, I think about our home completely differently. I just feel like we are at risk of our own property being damaged. All of our things. I can never be OK when it starts raining.

What could social research contribute to understanding Hurricane Ida's impact and improving the response—both to this disaster and those yet to come? Social research on previous hurricanes provides some examples. In 2017, Hurricane Harvey devastated Houston. Barrios et al. (2020) collected interview and participant observation data about the catastrophe among Houston's business elite, progressive residents of central historic neighborhoods, working-class African Americans in East and West Houston, and the predominantly Latino residents of the city's Manchester neighborhood. Hemmer and Elliff (2020) conducted interviews and focus groups with Houston-area school superintendents about their leadership experiences during the crisis. Other social researchers analyzed content in online databases regarding the reasons for and characteristics of school closures related to Hurricane Harvey (Jackson and Ahmed 2020). We use social science research like this on Hurricane Harvey and other disasters to illustrate how sociologists use qualitative methods to study the social world. These examples will help you

understand how some of our greatest insights into social processes can result from what appear to be very ordinary activities: observing, participating, listening, and talking.

But you will also learn that qualitative research is much more than just doing what comes naturally in social situations. Qualitative researchers must observe keenly, take notes systematically, question respondents strategically, and prepare to spend more time and invest more of their whole selves than often occurs with experiments or surveys. Moreover, if we are to have any confidence in a qualitative study's conclusions, each element of its design must be reviewed as carefully as we would review the elements of an experiment or survey. It's worth the effort since these methods can generate insights about the social world that may be overlooked with research relying solely on experiments or surveys.

The chapter begins with a brief history of qualitative methods and then an overview of the major features of qualitative research, using research on Hurricane Harvey and several other disasters for examples. The next section discusses the various approaches to participant observation research—the most distinctive qualitative method—and reviews the steps involved in participant observation. We also touch on the method of systematic observation, which is a quantitative approach to recording observational data; the contrast with participant observation will be instructive. Next, we review the method of intensive interviewing and then the method of focus groups. The last two sections discuss the challenge of generalizability in qualitative research and the ethical issues that are of concern in any type of qualitative research project. By the chapter's end, you should appreciate the hard work required to translate "doing what comes naturally" into systematic research, be able to recognize strong and weak points in qualitative studies, and be ready to do some of it yourself.

FUNDAMENTALS OF QUALITATIVE METHODS

Like quantitative methods, the term *qualitative methods* refers to a variety of research techniques that share some basic features. In this section, we will discuss the development of these methods and the features they have in common, and we provide details about several specific qualitative methods.

History of Qualitative Research

Anthropologists and sociologists laid the foundation for modern qualitative methods while doing **field research** in the early decades of the 20th century. Anthropologists Franz Boas and Bronislaw Malinowski, dissatisfied with studies of native peoples that relied on secondhand accounts and inspection of artifacts, went to live in or near the communities they studied. Boas visited Native American villages in the American Northwest; Malinowski lived among New Guinea natives. Neither truly participated in the ongoing social life of those they studied (Boas collected artifacts and original texts, and Malinowski reputedly lived as something of a noble among the natives he studied), but both helped establish the value of intimate familiarity with a community of interest (Emerson 1983:2–5).

Over time, many sociologists in the Department of Sociology and Anthropology at the University of Chicago adapted the fieldwork methods of anthropology to study the "natural areas" of cities and towns (Vidich and Lyman 2004). In the 1930s, 1940s, and 1950s, qualitative researchers were emphasizing the value of direct participation in community life and sharing in

subjects' perceptions and interpretations of events (Emerson 1983:6–13). This naturalistic focus continued to dominate qualitative research into the 1960s, and qualitative researchers refined and formalized methods to develop the most realistic understanding of the natural social world (Denzin and Lincoln 2000:14–15).

In the late twentieth century, many qualitative researchers focused more attention on the constructed nature of reality, meaning the belief that social reality is subjective and each person understands the world differently and in ways that can evolve over time (Saldana and Omasta 2021:128; Spretnak 1991:13–14). Some adherents urged qualitative researchers to view their interpretations of the social world as no more "privileged" than others' interpretations (Denzin and Lincoln 2000:15). In the 21st century, qualitative researchers are developing new techniques for studying social life online and for taking advantage of the increasing availability of pictures, videos, and texts on the Internet.

Features of Qualitative Research

Qualitative research designs share several features that distinguish them from experimental and survey research designs (Guba and Lincoln 1994; Maxwell 2005; Wolcott 1995):

Collection Primarily of Qualitative Rather Than Quantitative Data

Any research design may collect both qualitative and quantitative data, but qualitative methods emphasize observations that capture social life as participants experience it rather than in categories the researcher predetermines. For example, researchers at the Delaware Disaster Research Center observed the response to the unprecedented threat posed by Hurricane Katrina in the major New Orleans hotels and concluded the hotels went through three stages of "improvisation": (1) the hotels encouraged all guests who were not stranded to leave, (2) the hotel chains sent in food and other provisions for guests and staff, and (3) the hotels reorganized so they could provide semipermanent lodging for federal disaster employees and evacuees (Rodríguez, Trainor, and Quarantelli 2006:87–89). The researchers described the different activities in the hotels at each of these "stages."

A Focus on Previously Unstudied Processes and Unanticipated Phenomena

Previously unstudied attitudes and actions can't adequately be understood with a structured set of questions or within a highly controlled experiment. So qualitative methods have their greatest appeal when we need to explore new issues, investigate hard-to-study groups, or determine the meaning people give to their lives and actions. Disasters such as Hurricane Harvey certainly meet the criteria of unanticipated and hard to study. Hemmer and Elliff (2020) used qualitative methods to study how superintendents understood the conditions of their leadership and decisions made before, during, and after Hurricane Harvey because limited research existed on the experiences of superintendents who lead during times of crisis.

Exploratory Research Questions, With a Commitment to Inductive Reasoning

Qualitative researchers typically begin their projects seeking not to test preformulated hypotheses but to discover what people think, how they act, and why, in some social setting. Hemmer and Elliff (2020) began their research asking two questions: (1) How do superintendents make sense of their world and experiences during a crisis? and (2) How do superintendents describe actions made in the context of a crisis? Only after many observations do qualitative researchers try to develop general principles to account for their observations.

Sensitivity to the Subjective Role of the Researcher (Reflexivity)

Qualitative researchers recognize that their perspective on social phenomena will reflect in part their own background and current situation. Who the researcher is and "where they are coming from" can affect what the research "finds." Some qualitative researchers believe the goal of developing a purely "objective" view of the social world is impossible, but they discuss in their publications their own feelings about what they have studied so others can consider how these feelings affected their findings. Baker and Deham (2020) demonstrate **reflexivity** when they acknowledge their "lived experience of disaster both as citizen responders and researchers" informed their findings about citizen-based responses to Hurricane Harvey.

An Orientation to Social Context, to the Interconnections Between Social Phenomena Rather Than to Their Discrete Features

The context of concern may be a program or organization, a community, or a broader social context. This feature of qualitative research is evident in Franklin S. Allaire's (2020) description of how Hurricane Harvey impacted the experiences of first-semester college students, many of whom were first generation and members of underrepresented ethnic communities, at the University of Houston-Downtown.

> For the participants, there were several sources of Harvey-related stress and anxiety. First and foremost [for these first-semester college students] was the uncertainty that Harvey created concerning their classes and coursework. In their interviews, Ana, Cynthia, Guadalupe, Ivonne, Jay, Lee, Pam, and Yolanda explained how they wanted to be good students. However, they did not know what to do during these extraordinary circumstances. (p. 90)

A Focus on Human Subjectivity, on the Meanings That Participants Attach to Events and That People Give to Their Lives

"Through life stories, people 'account for their lives.'... The themes people create are the means by which they interpret and evaluate their life experiences and attempt to integrate these experiences to form a self-concept" (Kaufman 1986:24–25). You can see this emphasis in a personal communication Dworznik-Hoak (2020) had with a journalist who covered Hurricane Harvey:

> It sucks, but it wasn't the worst thing. The important thing was I didn't lose any loved ones during it. I lost possessions or materials, sure, but that's nothing really. Not like the stuff that matters. (Personal communication, October 2017) (p. 166)

Use of Idiographic Rather Than Nomothetic Causal Explanation

With its focus on particular actors and situations and the processes that connect them, qualitative research tends to identify causes as particular events embedded within an unfolding, interconnected action sequence (Maxwell 2005). The language of variables and hypotheses appears only rarely in the qualitative literature. In their analysis of the devastating 2005 Hurricane Katrina in New Orleans, Rodríguez et al. (2006) describe "emergent and prosocial behavior following Hurricane Katrina" by relaying the following sequence of events in New Orleans hospitals:

> The floodwaters from the levee breaks created a new kind of crisis. Basements with stored food, water, and fuel, as well as morgues, were inundated. . . . As emergency generators ran out of fuel, the water, sewage, and air-conditioning systems failed. Patients who died in the hospitals had to be temporarily stored in stairwells. Eventually, waste of

all kinds was strewn almost everywhere. The rising temperatures made most diagnostic equipment inoperable. . . . Regular hospital procedures simply stopped, but personnel improvised to try to provide at least minimum health care. For instance, physicians, nurses, and volunteers fanned patients to keep them cool, sometimes using manually operated devices to keep them breathing. (pp. 89–90)

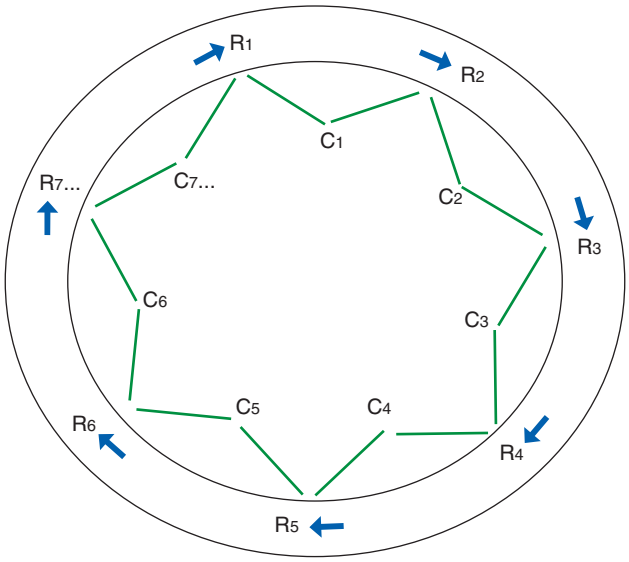
Acceptance—by Some Qualitative Researchers—of a Constructivist Philosophy

Constructivist social scientists believe social reality is socially constructed and the goal of social scientists is to understand what meanings people give to reality, not to determine how reality works apart from these constructions. This philosophy rejects the positivist belief that there is a concrete, objective reality that scientific methods help us understand (Fetterman 2019:7). Constructivist researchers believe "objects and events are understood by different people differently, and those perceptions are the reality—or realities—that social science should focus on" (Rubin and Rubin 1995:35).

Constructivists give particular attention to the different goals of researchers and other participants in a research setting and may seek to develop a consensus among participants about how to understand the focus of inquiry (Sulkunen 2008:73): "Truth is a matter of the best-informed and most sophisticated construction on which there is consensus at a given time" (Schwandt 1994:128).

Constructivist inquiry often involves an interactive research approach, in which a researcher begins an evaluation in some social setting by identifying the different interest groups in that setting. In a circular process known as a **hermeneutic circle** (Figure 10.1), the researcher interviews respondents representing these groups (R1, R2, etc.) to learn how they "construct" their thoughts

FIGURE 10.1 ■ The Hermeneutic Circle



Source: Guba and Lincoln, 1989. *Fourth Generation Evaluation.* SAGE.

and feelings about the topic of concern (C1, C2, etc.) and then gradually tries to develop a shared perspective on the problem being evaluated (Guba, Lincoln, and Lynham 2017).

Adaptive Research Design, in Which the Design Develops as the Research Progresses

An **adaptive research design** is one in which each component of the design may need to be reconsidered or modified in response to new developments or to changes in some other component. The activities of collecting and analyzing data, developing and modifying theory, elaborating or refocusing the research questions, and identifying and eliminating validity threats are usually all going on more or less simultaneously, each influencing all of the others (Maxwell 2005:2–3). You can see this adaptive quality in Levine's (2021) description of his qualitative research on the community development activities within the "Fairmont Corridor" in Boston, Massachusetts:

> In 2010, I conducted eight exploratory interviews to get a better sense of the Fairmount Line's history and the ongoing work by the coalition of nonprofit developers. After one of those interviews, the director of a community development corporation (CDC) invited me to attend a private coalition meeting. I was then invited to another, and then another. Meanwhile, I was soon attending all public meetings about development projects in the area. And so my fieldwork began. (p. 208)

STYLES OF QUALITATIVE RESEARCH

Case study and ethnographic research methods are designed to produce insights about whole settings and cultures. We will introduce both of these approaches and also show how the methods of digital ethnography and hybrid ethnography allow these research styles to be extended to social life conducted on the Internet.

The Case Study

Qualitative research projects often have the goal of developing an understanding of an entire slice of the social world, not just discrete parts of it. What was the larger social context in New Orleans after Hurricane Katrina (Rodríguez et al. 2006:87)? Sociologist Kai Erikson sent me the following verbal "picture" of New Orleans, as he observed it during a research trip a few days after Katrina:

> The carnage stretches out almost endlessly: more than a hundred thousand [crumpled] homes, at least fifty thousand [flattened] automobiles, the whole mass being covered by a crust of grey mud, dried as hard as fired clay by the sun. It was the silence of it, the emptiness of it; that is the story.

Questions and images such as these reflect a concern with developing a **case study**. Case study is not so much a single method as it is a way of thinking about what a qualitative research project can or perhaps should focus on. The case may be an organization, a community, a social group, a family, or even an individual; as far as the qualitative researcher is concerned, it must be understood in its specific social context (Tight 2017:19–20). The idea is that the social world functions as an integrated whole; social researchers therefore need to develop a "deep understanding of particular instances of phenomena" (Mabry 2008:214).

Central to much qualitative case study research is the goal of creating a **thick description** of the setting studied—a description that provides a sense of what it is like to experience that

setting from the standpoint of the natural actors in that setting (Geertz 1973). Sarkar (2021) illustrates how a thick description of life in Mumbai, India, during COVID-19 restrictions gives a feel of the place and persons within it:

> I wait in a queue outside the local Patel shop, along with around twenty others, standing in circles marked on the pavement at two-meter gaps. Two attendants stand at the shop entrance, one checking our temperatures, the other pouring sanitizer on our hands. They allow only five customers to go inside at a time, so the wait is long, but we are assured of maximized safety. As we wait, beggar children come and ask for money, or some ice cream. They wear masks nowadays, like all of us do. (p. 821)

You will learn in the next sections how qualitative methodologists design research that can generate such thick descriptions of particular cases.

Ethnography

Ethnography is the study of a culture or cultures that a group of people share (Van Maanen 1995:4). Many qualitative researchers are guided by the tradition of ethnography, and most specific qualitative methods share some of its techniques. As a method, ethnography usually refers to the process by which a single investigator immerses themself in a group for a long time (often one or more years), gradually establishing trust and experiencing the social world as do the participants (Madden 2010:16). Ethnographic research can be called *naturalistic* because it seeks to describe and understand the natural social world as it is, in all its richness and detail. This goal is best achieved when an ethnographer is fluent in the local language and spends enough time immersed in the setting to know how people live, what they say about themselves, what they do, and what they value (Armstrong 2008:55; Fawcett and Pockett 2015:63–64).

Ethnographic analysis relies on the thoroughness and insight of the researcher to "tell us like it is" in the setting, as they experienced it.

Code of the Street, Elijah Anderson's (1999) award-winning study of Philadelphia's inner city describes his concern with being "as objective as possible" and using his training "to look for and to recognize underlying assumptions, [those of the ethnographer] and those of their subjects, and to try to override the former and uncover the latter" (p. 11).

Anderson's report on his related Jelly's Bar study illustrates how his understanding deepened as he became more socially integrated into the group. He thus became more successful at "blending the local knowledge one has learned with what we already know sociologically about such settings" (Anderson 2003:236):

> I engaged the denizens of the corner and wrote detailed field notes about my experiences, and from time to time I looked for patterns and relationships in my notes. In this way, an understanding of the setting came to me in time, especially as I participated more fully in the life of the corner and wrote my field notes about my experiences; as my notes accumulated and as I reviewed them occasionally and supplemented them with conceptual memos to myself, their meanings became more clear, while even more questions emerged. (Anderson 2003:224)

A good ethnography like Anderson's is possible only when the ethnographer learns the subtleties of expression used in a group and the multiple meanings that can be given to statements or acts (Armstrong 2008:60–62). Good ethnographies also include some reflection by the researcher on the influence their own background has had on research plans, as well as on the impact of the research in the setting (Madden 2010:22–23).

Digital Ethnography

Communities can refer not only to people in a common physical location but also to relationships that develop online. Online communities may be formed by persons with similar interests or backgrounds or to supplement relationships that emerge in the course of work or school or other ongoing social activities. Like communities of people who interact face-to-face, online communities can develop a culture and become sources of identification and attachment (Kozinets 2010:14–15). As with physical communities, researchers can study online communities through immersion in the group for an extended period. **Digital ethnography**, also termed *netnography, cyberethnography,* and *virtual ethnography* (James and Busher 2009:34–35), is the use of ethnographic methods to study online communities.

In some respects, digital ethnography is similar to traditional ethnography. The researcher prepares to enter the field by becoming familiar with online communities and their language and customs, formulating an exploratory research question about social processes or orientations in that setting and selecting an appropriate community to study. Unlike in-person ethnographies, digital ethnographies can focus on communities whose members are physically distant and dispersed. The selected community should be relevant to the research question, involve frequent communication among actively engaged members, and have a number of participants who, as a result, generate a rich body of textual data (Kozinets 2010:89).

The digital ethnographer's self-introduction should be clear and friendly. Robert Kozinets (2010:93) provides the following example written about the online discussion space alt.coffee:

> I've been lurking here for a while, studying online coffee culture on alt.coffee, learning a lot, and enjoying it very much. . . . I just wanted to pop out of lurker status to let you know I am here. . . . I will be wanting to quote some of the great posts that have appeared here, and I will contact the individuals by personal e-mail who posted them to ask their permission to quote them. I also will be making the document on coffee culture available to any interested members of the newsgroup for their perusal and comments—to make sure I get things right.

A digital ethnographer must keep both observational and reflective field notes but, unlike a traditional ethnographer, can return to review the original data—the posted text—long after it was produced. The data can then be coded, annotated with the researcher's interpretations, checked against new data to evaluate the persistence of social patterns, and used to develop a theory that is grounded in the data.

But are you feeling a bit uncomfortable about using the term *ethnography* to describe investigations of people interacting in cyberspace? Not even an emoji can convey the rich sensory impressions we obtain in face-to-face social interaction; interpersonal contact mediated by digital technology is not just the same as direct contact (Pink et al. 2016:3). Even while we prepare to investigate the online social world with the full array of methods that have proven useful in investigating other areas of the social world, we need to keep in mind the context of the larger social world in which the digital world develops (Pink et al. 2016:8–14).

Hybrid Ethnography

It is becoming increasingly rare for a field site to have only in-person interaction. To truly immerse oneself in a field site, it is often necessary to join online interactions as well. Hybrid approaches use both traditional and digital ethnographic methods to capture both dimensions of a social setting (Przybylski 2020).

Hybrid ethnographers often collect massive quantities of data for analysis. Getting and staying organized is therefore particularly important. Hybrid ethnographers should organize online and offline field work time with an activity log and follow a regular schedule for engaging in the online and in-person portions of their field sites (Przybylski 2020). With modern technology, including social media, many people are documenting their own lives by creating video, audio, photos, and commentary (Przybylski 2020). Hybrid ethnographers must decide which data is pertinent to their research questions and should be included in their analyses.

Many qualitative projects focus on more specific aspects of the social world than ethnographic or case study research, but all qualitative designs draw on one or more of three specific techniques: participant observation, intensive (in-depth) interviewing, and focus groups. Two or three of these techniques may be combined in the same research project.

PARTICIPANT OBSERVATION

Participant observation is a qualitative method in which natural social processes are studied as they happen. By observing people and interacting with them during their normal activities, participant observers seek to avoid the artificiality of experimental design and the unnatural structured questioning of survey research (Fetterman 2019; Koegel 1987:8). This method encourages consideration of the context in which social interaction occurs, of the complex and interconnected nature of social relations, and of the sequencing of events (Bogdewic 1999:49).

Researchers using participant observation can adopt one of several different roles in a setting (see Figure 10.2) (Gold 1958). As a **covert observer**, a researcher observes others without participating in social interaction and does not self-identify as a researcher. This role is often adopted for studies in public places where there is nothing unusual about someone sitting and observing others. However, in many settings, a qualitative researcher will function as a **complete observer** who is publicly defined as a researcher and does not participate in group activities.

These two relatively passive roles contrast with the role of a researcher who participates actively in the setting. If the researcher publicly acknowledges being a researcher but nonetheless participates in group activities, they can be termed an *overt participant*, or true **participant observer**. By contrast, a qualitative researcher is a **complete participant** (also known as a *covert participant*) when they act just like other group members and do not disclose their research role.

Choosing a Role

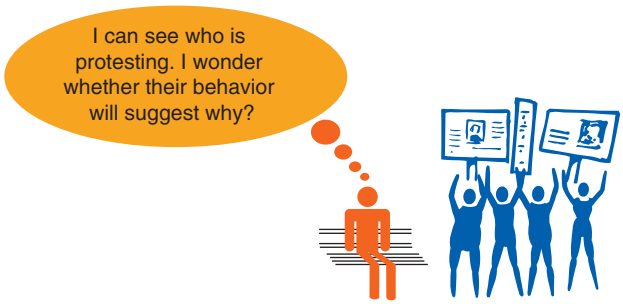
The first concern of every participant observer is thus to decide what balance to strike between observing and participating and whether to reveal one's role as a researcher. These decisions must consider the specifics of the social situation being studied, the researcher's own background and personality, the larger sociopolitical context, and ethical concerns. Which balance of participating and observing is most appropriate also changes during most projects—and often many times. Moreover, the researcher's ability to maintain either a covert or an overt role will many times be challenged. Although the specifics differ in online research, digital and hybrid ethnographers must also decide whether to announce their presence as researchers and how to participate in an online community and then consider how these decisions affect the ongoing online social interaction they are observing (Hewson et al. 2016:37; Pink et al. 2016:13–14).

Covert Observation

In both observational roles, researchers try to see things as they happen, without actively participating in these events. Although there is no fixed formula to guide the observational process,

FIGURE 10.2 ■ The Participant Observation Continuum

To study an activist group, you could take the role of a covert observer.



You could take the role of an overt observer:

You could take the role of a participant and observer:

You could take the role of a covert participant:

observers try to identify the who, what, when, where, why, and how of the activities in the setting. Their observations will usually become more focused over time, as the observer develops a sense of the important categories of people and activities and gradually develops a theory that accounts for what is observed (Fetterman 2019:48).

In social settings involving many people, in which observing while standing or sitting does not attract attention, covert observation is possible and is unlikely to have much effect on social processes. You may not even want to call this "covert" observation because your activities as an observer may be no different from those of others who are simply observing others to pass the

time. However, when you take notes, when you systematically check out the different areas of a public space or different people in a crowd, when you arrive and leave at particular times to do your observing, you are acting differently in important respects from others in the setting. Moreover, when you write up what you have observed and, possibly, publish it, you have taken something unique from the people in that setting. If you adopt the role of a covert observer, you should always remember to evaluate how your actions in the setting and your purposes for being there may affect the actions of others and your own interpretations.

Overt Observation

When a researcher announces their role as a research observer, their presence is much more likely to alter the social situation being observed. This is the problem of **reactive effects**. It is not "natural" in most social situations for someone to be present who will record their observations for research and publication purposes, and so individuals may alter their behavior. The overt, or complete, observer is even more likely to have an impact when the social setting involves few people or if observing is unlike the usual activities in the setting. Observable differences between the observer and those being observed also increase the likelihood of reactive effects. For example, in Parvez's (2018:461) research to understand the politics occurring among minority Muslims in secular states, Parvez engaged in open and warm conversations with low-income Muslim women in Hyderabad, India, but these women always viewed her as a member of the elite. An unequal power dynamic permeated all relationships Parvez formed with these low-income Muslim women.

Overt Participation (Participant Observer)

Most **field researchers** adopt a role that involves some active participation in the setting. Usually, they inform at least some group members of their research interests, but then they participate in enough group activities to develop rapport with members and to gain a direct sense of what group members experience. This is not an easy balancing act.

Unlike in India, where she limited her role to observing, Parvez (2018:461–462) functioned as a participant observer when she conducted her research on Muslims in Lyon, France:

> In Lyon, I was a participant-observer mostly in mosque communities […] I attended the women's Islamic lessons every week in Les Minguettes and spent additional hours in informal religious gatherings and Quranic recitation circles in people's apartments. Most of the women I came to know wore a hijab or djelbab (that covers all but the face) and faced daily discrimination over their veiling. Alongside them, I wore a hijab though not a djelbab. . . . I also witnessed them facing aggression from strangers specifically in public transit, which we used to commute to and from the mosque. Unlike in India, my French informants did not see me, as a woman of color, as wielding any obvious power over them due to class position or nationality. Indeed, several exoticized my identity, which I believe made them more willing to trust me.

Of course, the argument that the researcher's role can be disclosed without affecting the social process under investigation is less persuasive when the behavior to be observed is illegal or stigmatized so that participants have reasons to fear the consequences of disclosure to any outsider. Konstantin Belousov et al. (2007) provide a dramatic example of this problem from their fieldwork on regulatory enforcement in the Russian shipping industry. In a setting normally closed to outsiders and linked to organized crime, the permission of a port official was required. However, this official was murdered shortly after the research began. After that "our presence

was now barely tolerated, and to be avoided at all costs.... Explanations became short and respondents clearly wished to get rid of us as soon as possible" (pp. 164–165).

Even when researchers maintain a public identity as researchers, ethical dilemmas arising from participation in group activities do not go away. In fact, researchers may have to "prove themselves" to the group members by joining in some of their activities. Dum (2016) encountered this in a residential hotel that housed vulnerable populations (e.g., released prisoners, people with mental illness): "Drug and alcohol use were a key part of the motel's culture. I forged inroads with residents early on by sharing beer, and I was told on several occasions that establishing relationships at the motel was always easier when alcohol was involved" (p. 87).

Experienced participant observers try to lessen some of the problems of identity disclosure by evaluating both their effect on others in the setting and the effect of others on the observers writing about these effects throughout the time they are in the field and while they analyze their data. They also are sure, while in the field, to preserve some physical space and regular time when they can concentrate on their research and schedule occasional meetings with other researchers to review the fieldwork. Participant observers modify their role as circumstances seem to require, perhaps not always disclosing their research role at casual social gatherings or group outings but being sure to inform new members of it.

Covert Participation

To lessen the potential for reactive effects and to gain entry to otherwise inaccessible settings, some field researchers have adopted the role of covert participants, keeping their research secret and trying their best to act like other participants in a social setting or group. Laud Humphreys (1970) took the role of a covert participant when he served as a "watch queen" so he could learn about the men engaging in homosexual acts in a public restroom. Erving Goffman (1961) worked as a state hospital assistant while studying the treatment of psychiatric patients. These are two examples of covert participation.

RESEARCH IN THE NEWS

Community Collaboration After 1964 Anchorage Earthquake

In 1964, Anchorage, Alaska, experienced a 9.2 magnitude earthquake—the most powerful earthquake in American history. Social scientists from the newly founded Disaster Research Center (DRC) at Ohio State University (the DRC has since moved to the University of Delaware) traveled to Anchorage, Alaska, immediately after the earthquake for one week and returned five more times to the city over the next eighteen months, completing nearly 500 interviews that documented the community's response to the natural disaster. Although the U.S. government, which funded the DRC, predicted pandemonium and barbarism in response to disasters, the social scientists found an Anchorage community that was collaborative and compassionate. For example, Moallem (2020) writes, "When Presbyterian Hospital started filling with gas after the quake, Boy Scouts who'd been distributing phone books in the neighborhood helped walk the hospital's 22 patients down three or four flights of stairs, and an armada of taxis and other civilian drivers pulled up outside to evacuate everyone to a second hospital, across town." Since then, disaster research has consistently shown that catastrophes bring out the best in people.

A journalist's story can provide some of the richness of a qualitative case study and the insights that come from ethnographic immersion in a social context, but it's not the same as the approach of DRC social scientists and other systematic approaches to social research.

For Further Thought?

1. How well do you understand the social context of the disaster experience from the quoted description above? What else would you like to know? What questions would you like to ask survivors, and how would you select them?
2. What opportunities for conducting an ethnographic investigation can you think of in a disaster and recovery situation like the Anchorage earthquake? What would be some of the problems, even if you were living there and were familiar with the city?

News source: Mooallem, Jon. 2020. "This Is How You Live When the World Falls Apart." *The New York Times*, March 12:SR4.

Although the role of a covert participant lessens some of the reactive effects encountered by the complete observer, covert participants confront other problems:

- *Covert participants cannot take notes openly or use any obvious recording devices.* They must write up notes based solely on their memory and must do so at times when it is natural for them to be away from the group members.
- *Covert participants cannot ask questions that will arouse suspicion.* Thus, they often have trouble clarifying the meaning of other participants' attitudes or actions.
- *The role of a covert participant is difficult to play successfully.* Covert participants will not know how regular participants would act in every situation in which the researchers find themselves. Regular participants have entered the situation from different social backgrounds and with goals different from that of the researchers. Researchers' spontaneous reactions to every event are unlikely to be consistent with those of the regular participants (Mitchell 1993). Suspicion that researchers are not "one of us" may then have reactive effects, obviating the value of complete participation (Erikson 1967).
- *Covert participants need to keep up the act at all times while in the setting under study.* Researchers may experience enormous psychological strain, particularly in situations where they are expected to choose sides in intragroup conflict or to participate in criminal or other acts. Of course, some covert observers may become so wrapped up in the role they are playing that they adopt not only the mannerisms but also the perspectives and goals of the regular participants—that is, they "go native." At this point, they abandon research goals and cease to evaluate critically what they are observing.

Ethical issues have been at the forefront of debate over the strategy of covert participation. Erikson (1967) argued that covert participation is, by its very nature, unethical and should not be allowed except in public settings. Erikson points out that covert researchers cannot anticipate the unintended consequences of their actions for research subjects. In addition, other social scientists are harmed either when covert research is disclosed during the research or on its publication because distrust of social scientists increases and access to research opportunities may decrease. However, a total ban on covert participation would "kill many a project stone dead" (Punch 1994:90). Studies of unusual religious or sexual practices and of institutional malpractice would rarely be possible. Therefore, some field researchers argue that covert participation is legitimate in some settings. If the researcher maintains the confidentiality of others, keeps

commitments to others, and does not directly lie to others, some degree of deception may be justified in exchange for the knowledge gained (Punch 1994:90).

Entering the Field

Entering the field, the setting under investigation, is a critical stage in a participant observation project because it can shape many subsequent experiences. Some background work is necessary before entering the field—at least enough to develop a clear understanding of what the research questions are likely to be and to review one's personal stance toward the people and problems likely to be encountered. You need to have a sense of the social boundaries around the setting or "case" you will study. With participant observation, researchers must also learn in advance how participants dress and what their typical activities are to avoid being caught completely unaware.

Many field researchers avoid systematic study and extensive reading about a setting for fear it will bias their first impressions, but entering without any sense of the social norms can lead to a disaster. In *Street Corner Society*, Whyte (1955) describes what happened when he went to a hotel bar, without a good sense of social norms in his field site, in search of women to talk to:

> I looked around me again and now noticed a threesome: one man and two women. It occurred to me that here was a maldistribution of females which I might be able to rectify. I approached the group and opened with something like this: "Pardon me. Would you mind if I joined you?" There was a moment of silence while the man stared at me. He then offered to throw me downstairs. I assured him that this would not be necessary and demonstrated as much by walking right out of there without any assistance. (p. 289)

Whyte needed a **gatekeeper** who could grant him access to the setting; he finally found one in "Doc" (Rossman and Rallis 1998:108–111). A helpful social worker at the local settlement house introduced Whyte to this respected leader, who agreed to help:

> Well, any nights you want to see anything, I'll take you around. I can take you to the joints— gambling joints—I can take you around to the street corners. Just remember that you're my friend. That's all they need to know [so they won't bother you]. (Whyte 1955:291)

In short, field researchers must be very sensitive to the impression they make and to the ties they establish when entering the field. This stage lays the groundwork for collecting data from people who have different perspectives and for developing relationships the researcher can use to surmount the problems in data collection that inevitably arise in the field. The researcher should be ready with a rationale for their participation and some sense of the potential benefits to participants. Discussion about these issues with key participants or gatekeepers should be honest and identify what the participants can expect from the research, without necessarily going into detail about the researcher's hypotheses or research questions (Rossman and Rallis 1998:51–53, 105–108).

Developing and Maintaining Relationships

Researchers must be careful to manage their relationships in the research setting so they can continue to observe and interview diverse members of the social setting throughout the long period typical of participant observation (Maxwell 2005:82–87). Every action the researcher takes can develop or undermine this relationship. Interaction early in the research process is particularly sensitive because participants don't know the researcher and the researcher doesn't know the

participants. Van Maanen (1982) found his relationship with police officers undermined by one incident:

> Following a family beef call in what was tagged the Little Africa section of town, I once got into what I regarded as a soft but nonetheless heated debate with the officer I was working with that evening on the merits of residential desegregation. My more or less liberal leanings on the matter were bothersome to this officer, who later reported my disturbing thoughts to his friends in the squad. Before long, I was an anathema to this friendship clique and labeled by them undesirable. Members of this group refused to work with me again. (p. 110)

Do you think Van Maanen should have kept his opinions about residential desegregation to himself? How honest should field researchers be about their feelings? Should they "go along to get along"?

Experienced participant observers have developed some sound advice for others seeking to maintain relationships in the field (Bogdewic 1999:53–54; Rossman and Rallis 1998:105–108; Whyte 1955:300–306; Wolcott 1995:91–95):

- Develop a plausible (and honest) explanation for yourself and your study.
- Maintain the support of key individuals in groups or organizations under study.
- Be unobtrusive and unassuming. Don't "show off" your expertise.
- Don't be too aggressive in questioning others (e.g., don't violate implicit norms that preclude discussion of illegal activity with outsiders).
- Ask very sensitive questions only of informants with whom your relationship is good.
- Be self-revealing but only up to a point. Let participants learn about you as a person but without making too much of yourself.
- Don't fake your social similarity with your subjects. Taking a friendly interest in them should be an adequate basis for developing trust.
- Avoid giving or receiving monetary or other tangible gifts but without violating norms of reciprocity. Such small forms of assistance as an occasional ride to the store or advice on applying to college may strike the right balance.
- Be prepared for special difficulties and tensions if multiple groups are involved. It is hard to avoid taking sides or being used in situations of intergroup conflict.

Sampling People and Events

In qualitative research, the need to intensively study the people, places, or phenomena of interest guide sampling decisions. Most qualitative researchers limit their focus to just one or a few sites, so they can focus all their attention on the social dynamics of those settings. This does not mean that sampling is unimportant. The researcher must be reasonably confident about gaining access and that the site can provide relevant information. The sample must be appropriate and adequate for the study, even if it is not representative. The qualitative researcher may select a *critical case* that is unusually rich in information pertaining to the research question, a *typical case* precisely because it is judged to be typical, or a *deviant case* that provides a useful contrast (Kuzel 1999). Within a research site, plans may be made to sample different settings, people, events, and artifacts (see Table 10.1).

TABLE 10.1 ■ Sampling Plan for a Participant Observation Project in Schools

Information Source*			Type of Information to Be Obtained		
	Collegiality	Goals and Community	Action Expectations	Knowledge Orientation	Base
Settings					
Public places (halls, main offices)					
Teacher's lounge	X	X		X	X
Classrooms		X	X	X	X
Meeting rooms	X		X	X	
Gymnasium or locker room		X			
Events					
Faculty meetings	X		X		X
Lunch hour	X[1]				X
Teaching		X	X	X	X
People					
Principal		X	X	X	X
Teachers	X*	X	X	X	X
Students		X	X	X	
Artifacts					
Newspapers		X	X		X
Decorations		X*			

Source: Adapted from Marshall and Rossman (2011:108–110). Reprinted with permission from SAGE Publications, Inc.

1 *Selected examples in each category.

Studying more than one case or setting almost always strengthens the causal conclusions and makes the findings more generalizable (King, Keohane, and Verba 1994). For example, Barrios et al. (2020) engaged residents from specific Houston neighborhoods (West Houston, East Houston, Manchester, 5th Ward, 2nd Ward, Forest Cove, and the central historic neighborhoods directly west of the city's downtown area) to explore the meaning made of Hurricane Harvey among Houstonians of various socioeconomic backgrounds and political orientations.

Other approaches to sampling in field research are more systematic. You have already learned in Chapter 5 about some of the nonprobability sampling methods used in field research. For instance, purposive sampling can be used to identify the opinion of the leaders and representatives of different roles. With snowball sampling, field researchers learn from participants about who represents different subgroups in a setting. Quota sampling also may be employed to ensure the representation of particular categories of participants. Using some type of intentional sampling strategy within a particular setting can allow tests of some hypotheses, which would otherwise have to wait until comparative data could be collected from several other settings (King et al. 1994).

Theoretical sampling (see Figure 10.3) is a systematic approach to sampling in participant observation studies (Glaser and Strauss 1967). Saldana and Omasta (2021:312) define theoretical sampling as "the search for participants who might be able to confirm or disconfirm the patterns and theories developed by the researcher to date; particularly useful in the development of grounded theory." In her digital ethnography exploring the meaning of "popularity" in the YouTube beauty community, García-Rapp (2019) completed an initial tagging of all 313 videos uploaded by a beauty YouTuber and then went on to strategically select videos for further analysis to allow for exploration of questions that arose from the initial tagging.

FIGURE 10.3 ■ Theoretical Sampling

Original cases interviewed in a study of cocaine users:
Realization: Some cocaine users are businesspeople.
Add businesspeople to sample:
+
Realization: Sample is low on women.
Add women to sample:
+
Realization: Some female cocaine users are mothers of young children.
Add mothers to sample:
+

Taking Notes

Written notes are the primary means of recording participant observation data (Emerson, Fretz, and Shaw 1995). Many field researchers jot down partial notes while observing and then retreat to their computers to write up more complete notes daily.

It is almost always a mistake to try to take comprehensive notes while engaged in the field—the process of writing extensively is just too disruptive. The usual procedure is to jot down brief notes about highlights of the observation period. These brief notes, called **jottings**, can then serve as memory joggers when the researcher is writing the actual **field notes** at a later session. With the aid of the jottings and some practice, researchers usually remember a great deal of what happened—as long as the comprehensive field notes are written immediately afterward—or at least within the next 24 hours—and before they have discussed them with anyone else.

Usually, writing up notes takes much longer—at least three times longer—than the observing. Field notes must be as complete, detailed, and true as possible to what was observed and heard. Direct quotes should be distinguished clearly from paraphrased quotes, and both should be set off from the researcher's observation and reflections. Pauses and interruptions should be indicated. The surrounding context should receive as much attention as possible, and a map of the setting should always be included with indications of where the individuals were at different times.

Careful note-taking yields a big payoff. On page after page, field notes will suggest new concepts, causal connections, and theoretical propositions. Social processes and settings can be described in rich detail, with ample illustrations. Figure 10.4, for example, contains field notes recorded by Norma Ware, an anthropologist studying living arrangements for homeless mentally ill persons in the Boston housing study for which I was a coinvestigator (Schutt 2011b). The notes contain observations of the setting, the questions the anthropologist asked, the answers she received, and her analytic thoughts about one of the residents.

Complete field notes must provide even more than a record of what was observed or heard. Notes also should include descriptions of the methodology: where researchers were standing or sitting while they observed, how they chose people for conversation or observation, what counts of people or events they made, and why. Sprinkled throughout the notes also should be a record of the researchers' feelings and thoughts while observing: when they were disgusted by some statement or act, when they felt threatened or intimidated, why their attention shifted from one group to another, and what ethical concerns arose. Notes such as these provide a foundation for later review of the likelihood of bias or of inattention to some salient features of the situation.

Notes may, in some situations, be supplemented by still pictures, videotapes, and printed material circulated or posted in the research setting. Such visual material can bring an entirely different qualitative dimension into the analysis and call attention to some features of the social situation and actors within it that were missed in the notes (Grady 1996). Commentary on this material can be integrated with the written notes (Bogdewic 1999:67–68).

Managing the Personal Dimensions

Our overview of participant observation would not be complete without considering its personal dimensions. Because field researchers become a part of the social situation they are studying, they cannot help but be affected on a personal, emotional level. At the same time, those being studied react to researchers not just as researchers but as personal acquaintances—often as friends, sometimes as personal rivals. Managing and learning from this personal side of field research is an important part of any project.

FIGURE 10.4 ■ Field Notes From an Evolving Consumer Household (ECH)

I arrive around 4:30 p.m. and walk into a conversation between Jim and somebody else as to what color jeans he should buy. There is quite a lot of joking going on between Jim and Susan. I go out to the kitchen and find Dick about to take his dinner out to the picnic table to eat (his idea?) so I go ask if I can join him. He says yes. In the course of the conversation, I find out that he works 3 days a week in the "prevoc" program at the local day program, Food Services branch, for which he gets $10 per week. Does he think the living situation will work out? Yes. All they need is a plan for things like when somebody buys something and then everybody else uses it. Like he bought a gallon of milk and it was gone in two days, because everyone was using it for their coffee. I ask if he's gone back to the shelter to visit and he says, "No. I was glad to get out of there." He came to [the ECH] from [a shelter] through homeless outreach [a Department of Mental Health program]. Had been at [the shelter] since January. Affirms that [the ECH] is a better place to live than the shelter. Why? Because you have your own room and privacy and stuff. How have people been getting along with each other? He says, "Fine."

I return to the living room and sit down on the couch with Jim and Susan. Susan teases Jim and he jokes back. Susan is eating a T.V. dinner with M & M's for dessert. There is joking about working off the calories from the M & M's by doing sit-ups, which she proceeds to demonstrate. This leads to a conversation about exercise during which Jim declares his intention to get back into exercise by doing sports, like basketball.

Jim seems to have his mind on pulling himself together, which he characterizes as "getting my old self back." When I ask him what he's been doing since I saw him last, he says, "Working on my appearance." And in fact, he has had a haircut, a shave, and washed his clothes. When I ask him what his old self was like, he says, "You mean before I lost everything?" I learn that he used to work two jobs, had "a family" and was into "religion." This seems to have been when he was quite young, around eighteen. He tells me he was on the street for 7–8 years, from 1978–1985, drinking the whole time. I ask him whether he thinks living at [the ECH] will help him to get his "old self back" and he says that it will "help motivate me." I observe that he seems pretty motivated already. He says yes, "but this will motivate me more."

Jim has a warm personality, likes to joke and laugh. He also speaks up—in meetings he is among the first to say what he thinks and he talks among the most. His "team" relationship with Bill is also important to him—"me and Bill, we work together."

Source: Field notes from an ECH made available by Norma Ware, unpublished ethnographic notes, 1991.

The impact of personal issues varies with the depth of researchers' involvement in the setting. The more involved researchers are in the multiple aspects of the ongoing social situation, the more important personal issues become and the greater the risk of "going native." Even when researchers acknowledge their role, "increased contact brings sympathy, and sympathy in its turn dulls the edge of criticism" (Fenno 1978:277). There is no formula for successfully managing the personal dimension of a field research project. It is much more an art than a science and flows more from the researcher's own personality and natural approach to other people than from formal training. Sharing similarities such as age, race, or gender with those who are studied may help create mutual feelings of comfort, but such social similarities may mask more important differences in perspective resulting from education, social class, and having the role of researcher (Doucet and Mauthner, 2008:334).

Novice field researchers often neglect to consider how they will manage personal relationships when they plan and carry out their projects. Then, suddenly, they find themselves doing something they don't believe they should just to stay in the good graces of research subjects or juggling the emotions resulting from conflict within the group. As Whyte (1955) noted,

> The field worker cannot afford to think only of learning to live with others in the field. He has to continue living with himself. If the participant observer finds himself engaging in behavior that he has learned to think of as immoral, then he is likely to begin to wonder what sort of a person he is after all. Unless the field worker can carry with him a reasonably consistent picture of himself, he is likely to run into difficulties. (p. 317)

If you plan a field research project, follow these guidelines:

- Take the time to consider how you want to relate to your potential subjects as people.
- Speculate about what personal problems might arise and how you will respond to them.
- Keep in touch with other researchers and personal friends outside the research setting.
- Maintain standards of conduct that make you comfortable as a person and that respect the integrity of your subjects. (Whyte 1955:300–317)

INTENSIVE INTERVIEWING

Intensive (in-depth) interviewing is a qualitative method of finding out about people's experiences, thoughts, and feelings. Although intensive interviewing can be an important element in a participant observation study, it is often used by itself (Wolcott 1995:102–105). It shares with other qualitative research methods a commitment to learning about people in-depth and on their own terms and in the context of their situations.

Unlike the more structured interviewing that may be used in survey research (discussed in Chapter 8), intensive or in-depth interviewing relies on open-ended questions. Rather than asking standard questions in a fixed order, intensive interviewers may allow the specific content and order of questions to vary from one interviewee to another. Rather than presenting fixed responses that presume awareness of the range of answers respondents might give, intensive interviewers expect respondents to answer questions in their own words.

What distinguishes intensive interviewing from less structured forms of questioning is consistency and thoroughness. The goal is to develop a comprehensive picture of the interviewee's background, attitudes, and actions, in their own terms—that is, to "listen to people as they describe how they understand the worlds in which they live and work" (Rubin and Rubin 1995:3). For example, Dworznik-Hoak (2020) sought through in-depth interviewing of reporters who covered Hurricane Harvey to learn how these journalists "describe their experiences covering Harvey including triumphs and hardships, emotions felt, and reactions experienced, as well as significant moments" (p. 164).

Intensive interview studies do not reveal as directly as does participant observation the social context in which action is taken and opinions are formed. Nonetheless, in-depth interviewers seek to account for context. Jack Douglas (1985) made the point succinctly in *Creative Interviewing:*

> *Creative interviewing is purposefully situated interviewing.* Rather than denying or failing to see the situation of the interview as a determinant of what goes in the questioning and answering processes, creative interviewing embraces the immediate, concrete situation; tries to understand how it is affecting what is communicated; and, by understanding these effects, changes the interviewer's communication processes to increase the discovery of the truth about human beings. (p. 22)

Similar to participant observation studies, intensive interviewing engages researchers more actively with subjects than standard survey research does. The researchers must listen to lengthy explanations, ask follow-up questions tailored to the preceding answers, and seek to learn about interrelated belief systems or personal approaches to things rather than measure a limited set of variables. As a result, intensive interviews are often much longer than standardized interviews, sometimes as long as 15 hours, conducted in several different sessions. The intensive interview becomes more like a conversation between partners than an interview between a researcher and a subject (Kaufman 1986:22–23). Some call it "a conversation with a purpose" (Rossman and Rallis 1998:126).

Intensive interviewers actively try to probe understandings and engage interviewees in a dialogue about what they mean by their comments. To prepare for this active interviewing, the interviewer should learn in advance about the setting to be studied. Preliminary discussion with **key informants**, inspection of written documents, and even a review of your own feelings about the setting can all help (Miller and Crabtree 1999c:94–96). Allaire (2020) elaborates on this aspect of intensive interviewing when describing the methods he used in his study of first-year college students at the University of Houston-Downtown in the aftermath of Hurricane Harvey. Allaire (2020) ensured the themes identified in his research were an accurate representation of what the students discussed during interviews by having student participants read and review their interview transcripts, using multiple data sources to provide multiple perspectives on a given event, and recruiting a colleague to review narratives to gauge the accuracy of identified themes. Allaire writes:

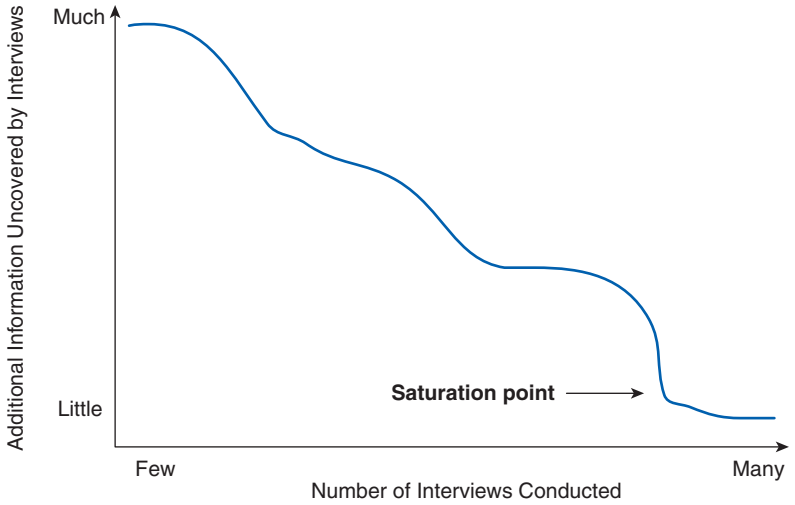
> As a resident of the Houston-metro area, I was also impacted by Hurricane Harvey. My experiences before, during, and after Hurricane Harvey mirror the themes explored here and sharing Harvey-related experiences built trust and rapport with the participants. However, this present study was about the participants' experiences and realities, not mine. As such, the methods mentioned earlier helped ensure that the themes represented the participants, their experiences, and their interpretation of their individual and collective realities without the infiltration of my own experiences and biases. (p. 89)

The intensive interview follows a preplanned outline of topics. It may begin with a few simple questions that gather background information while building rapport. These are often followed by a few general **grand tour questions** that are meant to elicit narratives to provide a broad overview of a social scenario from the participant's pointof view (Fetterman 2019). Some projects may use relatively structured interviews, particularly when the focus is on developing knowledge about prior events or some narrowly defined topic. But more exploratory projects, particularly those aiming to learn about interviewees' interpretations of the world, may let each interview flow in a unique direction in response to the interviewee's experiences and interests (Fetterman 2019; Kvale 1996:3–5; Rubin and Rubin 1995:6; Wolcott 1995:113–114). In either case, qualitative interviewers must adapt nimbly throughout the interview, paying attention to nonverbal cues, expressions with symbolic value, and the ebb and flow of the interviewee's feelings and interests. "You have to be free to follow your data where they lead" (Rubin and Rubin 1995:64).

Random selection is rarely used to select respondents for intensive interviews, but the selection method still must be considered carefully. If interviewees are selected in a haphazard manner, as by speaking just to those who happen to be available at the time when the researcher is on site, the interviews are likely to be of less value than when a more purposive selection strategy is used. Researchers should try to select interviewees who are knowledgeable about the subject of the interview, who are open to talking, and who represent the range of perspectives (Rubin and

Rubin 1995:65–92). Selection of new interviewees should continue, if possible, at least until the **saturation point** is reached, the point when new interviews seem to yield little additional information (see Figure 10.5). As new issues are uncovered, additional interviewees may be selected to represent different opinions about these issues.

FIGURE 10.5 ■ The Saturation Point in Intensive Interviewing



Establishing and Maintaining a Partnership

Because intensive interviewing does not engage researchers as participants in subjects' daily affairs, the problems of entering the field are much reduced. Regardless, it is important to establish rapport with subjects by considering in advance how they will react to the interview arrangements and by developing an approach that does not violate their standards for social behavior. Interviewees should be treated with respect, as knowledgeable partners whose time is valued (in other words, avoid coming late for appointments). A commitment to confidentiality should be stated and honored (Rubin and Rubin 1995).

In the first few minutes of the interview, the goal is to show interest in the interviewee and to explain clearly the purpose of the interview (p. 128). During the interview, the interviewer should maintain an appropriate distance from the interviewee, one that doesn't violate cultural norms; the interviewer should maintain eye contact—although this may not be appropriate in cultures in which eye contact is avoided—and not engage in distracting behavior. An appropriate pace is also important; pause to allow the interviewee to reflect, elaborate, and generally not feel rushed (Gordon 1992). When an interview covers emotional or otherwise stressful topics, the interviewer should give the interviewee an opportunity to unwind at the interview's end (Rubin and Rubin 1995:138). More generally, intensive interviewers must be sensitive to the broader social context of their interaction with the interviewee and to the implications of their relationship in the way they ask questions and interpret answers. It is also important to acknowledge that the intensive interviewer's relationship with the interviewee is not an equal partnership because the researcher seeks to gain information from the participant (Kvale 1996:6).

Asking Questions and Recording Answers

Intensive interviewers must plan their main questions around an outline of the interview topic. The questions should generally be short and to the point. More details can then be elicited through nondirective probes (e.g., "Can you tell me more about that?" or "uh-huh," echoing the respondent's comment, or just maintaining a moment of silence). Follow-up questions can then be tailored to answers to the main questions.

Interviewers should strategize throughout an interview about how best to achieve their objectives while accounting for interviewees' answers. *Habits of the Heart*, which explores the role of religion in people's lives in modern American society, provides a useful illustration (Bellah et al. 1985:304):

> [Coinvestigator Steven] Tipton, in interviewing Margaret Oldham [a pseudonym], tried to discover at what point she would take responsibility for another human being:
>
> **Q:** *So what are you responsible for?*
> **A:** *I'm responsible for my acts and for what I do.*
> **Q:** *Does that mean you're responsible for others, too?*
> **A:** *No.*
> **Q:** *Are you your sister's keeper?*
> **A:** *No.*
> **Q:** *Your brother's keeper?*
> **A:** *No.*
> **Q:** *Are you responsible for your husband?*
> **A:** *I'm not. He makes his own decisions. He is his own person. He acts his own acts. I can agree with them, or I can disagree with them. If I ever find them nauseous enough, I have a responsibility to leave and not deal with it any more.*
> **Q:** *What about children?*
> **A:** *I. . . . I would say I have a legal responsibility for them, but in a sense I think they in turn are responsible for their own acts.*

Do you see how the interviewer actively encouraged the subject to explain what she meant by "responsibility"? This sort of active questioning undoubtedly did a better job of clarifying her concept of responsibility than a fixed set of questions would have.

Interviewing Online

Our social world now includes many connections initiated and maintained through e-mail and other forms of web-based communication, so it is only natural that interviewing has also moved online. Online interviewing can facilitate interviews with others who are separated by physical distance; it also is a means to conduct research with those who are known only through such online connections as a discussion group, an e-mail distribution list, or social media (James and Busher 2009:14; Salmons 2012; Paulus and Lester 2021). As with digital ethnography, however, it is important to keep in mind how the online interview experience differs from face-to-face interviews and to consider the larger social context in which it takes place.

Online interviews can be either *synchronous*—in which the interviewer and interviewee exchange messages, as in online chatting or with text messages—or *asynchronous*—in which the interviewee can respond to the interviewer's questions whenever it is convenient, usually through e-mail but perhaps through a blog, a wiki, or an online forum (Salmons 2012; Paulus and Lester

2021). Both styles of online interviewing have advantages and disadvantages (James and Busher 2009:13–16). Synchronous interviewing provides an experience more like an in-person interview, thus giving more of a sense of obtaining spontaneous reactions, but it requires careful attention to arrangements and is prone to interruptions (Paulus and Lester 2021). Asynchronous interviewing allows interviewees to provide more thoughtful and developed answers, but it may be difficult to maintain interest and engagement if the exchanges continue over many days. The online asynchronous interviewer should plan carefully how to build rapport as well as how to terminate the online relationship after the interview is concluded (King and Horrocks 2010:86–93). Adding video to the exchange can also increase engagement, whether through real-time videoconferencing or by sending video clips or podcasts (Salmons 2012:4–5).

Gruber et al. (2021) initially planned to conduct face-to-face interview with irregular immigrants in several European countries but had to adapt once the coronavirus pandemic began. The researchers shifted to a mixed-mode design, given the pandemic conditions in migrants' local areas. Gruber et al. (2021) write "In concrete terms, this means that face-to-face interviews, remote video interviews, and very flexible forms of asynchronous text-based interviews will be applied, depending on the situation of the interviewees (i.e., whether they live in crowded reception centers or small private accommodations, whether they are literate, whether they are flexible regarding time or difficult to reach, etc.), the technical requirements, and the capacities of the people responsible in the centers" (p. 11). The researchers adopted a flexible approach, with both synchronous and asynchronous options, in order to be able to access a hard-to-reach population during a global pandemic.

Whether a synchronous or asynchronous approach is used, online interviewing can facilitate the research process by creating a written record of the entire interaction without the need for typed transcripts. The relative anonymity of online communications can also encourage interviewees to be more open and honest about their feelings than they would be if interviewed in person (James and Busher 2009:24–25). However, online interviewing lacks some of the most appealing elements of qualitative methods: The revealing subtleties of facial expression, intonation, and body language are lost unless video is also added, and the intimate rapport that a good intensive interviewer can develop in a face-to-face interview may not be achieved.

People may create different online personas, and this too becomes an important part of the social world to investigate—even if these online personas differ from the people who create them. Ann Randall (2012) wondered whether Second Life members would experience beneficial effects from being interviewed in Second Life through their avatars, in the same way that "real people" may report that participating in qualitative interviews provides benefits like feeling respected and having an opportunity to reflect on their experiences. Second Life is a three-dimensional virtual world in which real people are represented by avatars they create or purchase. To explore this issue, Randall invited Second Life members to complete qualitative interviews as their avatars. Avatars who agreed to participate returned a consent form with their avatar's name (retaining the anonymity of the interview). Of her small sample of avatars, three chose to be interviewed in their own Second Life homes while the other six came to virtual locations that Randall created for the purpose, such as a beach house with reclining chairs. Interviews were semi-structured, with a set of basic questions but the opportunity to pursue issues with additional questions as they arose.

The interview began with a friendly introductory exchange, like the following between Randall's avatar and the participant's avatar (Randall 2012:139):

Ann: *Hi!!*
Participant: *hello*

Ann: *nice to meet you*
Participant: *likewise*
Ann: *feel free to have a seat*
Participant: *thanks*
Ann: *you can sit on the stump if you want. I've restricted access so we can speak privately. Oh and have some coffee if you'd like. Right click on the coffee pot.*

Six months later, Randall scheduled interviews with the same nine participants to ask them about possible beneficial effects of the interview. She reports finding much evidence of such beneficial effects, just as if actual people had been interviewed in person.

FOCUS GROUPS

A **focus group** is "a 'non-standard' technique for information gathering based on an apparently informal discussion among a group of people selected on the basis of specific characteristics, outlined according to the cognitive purposes of the research" (Acocella and Cataldi 2020:4). A researcher is present who presents specific questions and guides the discussion, but the resulting information is qualitative and relatively unstructured. Focus group research projects usually involve a series of focus groups involving similar participants.

During focus groups, the researcher, or moderator, uses an interview guide, but the level of directivity employed by a moderator and the accompanying freedom of conversation can vary greatly (Acocella and Cataldi 2020:163). Focus group moderators must begin the discussion by generating interest in the topic, creating the expectation that all will participate, and making it clear that the researcher does not favor any particular perspective or participant (Smithson 2008:361). All questions should be clear, simple, and straightforward. The moderator should begin with easy-to-answer general factual questions and then, about one quarter to halfway through the allotted time, shift to key questions on specific issues. If the question flow is successful, the participants should experience the focus group as an engaging and friendly discussion and the moderator should spend more time after the introductory period listening and guiding the discussion than asking questions.

Most focus groups involve 5 to 10 people, a number that facilitates discussion by all in attendance (Krueger and Casey 2009:6). Participants usually do not know one another, although some studies in organized settings may include friends or coworkers. Opinions differ on the value of using homogeneous versus heterogeneous participants. Homogeneous groups may be more convivial and willing to share feelings, but heterogeneous groups may stimulate more ideas (Brown 1999:115–117). In any case, it is important to avoid having some participants who have supervisory or other forms of authority over other participants (Krueger and Casey 2009:22). It is also good to avoid focus groups when dealing with emotionally charged issues, when sensitive information is needed and when confidentiality cannot be ensured, or when the goal is to reach consensus (Krueger and Casey 2009:20).

Keith Elder and his colleagues at the University of South Carolina and elsewhere (Elder et al. 2007:S125) used focus groups to study the decisions by African Americans not to evacuate New Orleans before Hurricane Katrina. Elder et al. conducted six focus groups with 53 evacuees who were living in hotels in Columbia, South Carolina, between October 3 and October 14, 2005.

One of the themes identified in the focus groups was the confusion resulting from inconsistent messages about the storm's likely severity (Elder et al. 2007:S126):

> The mayor did not say it was a mandatory evacuation at first. One or two days before the hurricane hit, he said it was mandatory. It was too late then.

> They didn't give us no warning. . . . When they said leave, it was already too late.
>
> After [the] levees broke the mayor said mandatory evacuation, before then he was not saying mandatory evacuation.
>
> Governor said on TV, you didn't want to go, you didn't have to go, cause it was no threat to us, she said.

Careful planning of group moderation to encourage diverse opinions should be implemented. Disagreements should be handled carefully so no participants feel uncomfortable, and the discussion should be kept focused on the announced topic (Smithson 2008:361). Moderators can be trained to be "devil's advocates" and argue for contrary opinions or to support opinions proposed by a minority of participants to encourage the expression of alternative viewpoints (MacDougall and Baum, 1997; Acocella and Cataldi 2020:224). The moderator may conclude the group discussion by asking participants for recommendations to policy makers or their further thoughts that they have not had a chance to express (Krueger and Casey 2009:36–48). The issue of impact of interviewer style and questioning must be considered when evaluating the results of focus groups.

No formal procedure exists for determining the generalizability of focus group answers, but the careful researcher should conduct at least several focus groups on the same topic and check for consistency in the findings. When differences in attitudes between different types of people are a concern, separate focus groups may be conducted that include these different types, and then the analyst can compare comments between them (Krueger and Casey 2009:21). Some focus group experts advise conducting enough focus groups to reach the point of saturation, when an additional focus group adds little new information to that which already has been generated (Brown 1999:118). Because it is not possible to conduct focus groups with large, representative samples, it is always important to consider how recruitment procedures have shaped the generalizability of focus group findings.

GENERALIZABILITY IN QUALITATIVE RESEARCH

Qualitative research often focuses on populations that are hard to locate or are very limited in size. In consequence, nonprobability sampling methods such as availability sampling and snowball sampling are often used (see Chapter 5). However, this does not mean that generalizability should be ignored in qualitative research (Gobo 2008:206). Janet Ward Schofield (2002) suggests two different ways of increasing generalizability in qualitative investigations:

- *Studying the typical.* Choosing sites on the basis of their fit with a typical situation is far preferable to choosing on the basis of convenience. (p. 181)
- *Performing multisite studies.* A finding emerging repeatedly in the study of numerous sites would appear to be more likely to be a good working hypothesis about some as yet unstudied site than a finding emerging from just one or two sites.... Generally speaking, a finding emerging from the study of several very heterogeneous sites would be more... likely to be useful in understanding various other sites than one emerging from the study of several very similar sites. (p. 184)

Giampietro Gobo (2008:204–205) highlights another approach to improving generalizability in qualitative research. A case may be selected for in-depth study because it is atypical,

or deviant. Investigating social processes in a situation that differs from the norm will improve understanding of how social processes work in typical situations: "the exception that proves the rule."

At the same time, some qualitative researchers question the value of generalizability, as most researchers understand it. The argument is that understanding the particulars of a situation in depth is an important object of inquiry in itself. In the words of sociologist Norman Denzin,

> The interpretivist rejects generalization as a goal and never aims to draw randomly selected samples of human experience. . . . Every instance of social interaction . . . represents a slice from the life world that is the proper subject matter for interpretive inquiry. (Denzin, as cited in Schofield 2002:173)

You will have to decide for yourself whether it makes sense to be concerned with generalizability in qualitative research.

ETHICAL ISSUES

When qualitative researchers engage actively in the social world—in other words, when they are doing their job—they encounter unique ethical challenges.

It is important to note that the natural, evolving quality of much interaction between qualitative researchers and participants can make it difficult in many situations to stipulate in advance the procedures that will be followed to ensure ethical practice. In the words of Bosk and De Vries (2004),

> Few of us start with specific hypotheses that we will later test in any systematic way. . . . We cannot state our procedures any more formally than we will hang around here in this particular neighborhood and try to figure out what is going on among these people. (P. 253)

Just as in everyday social interaction, a qualitative researcher's engagement with others "in the field" can go in a direction that was not anticipated in the formal research plan. Or as Martin Levinson (2010:195) put it, "the very nature of much ethnographic work—its openness and uncertainty—is such that a cunning researcher can easily circumvent ethical fences that are nominally in place." As an example, Levinson (2010:196–197) recounts his effort to gain the consent of a potential interviewee in his ethnographic study of Romani people in the United Kingdom:

> **Me [Levinson]:** *So, would you mind answering a few questions?*
> *No response.*
> **Me:** *Well, is there anything you'd like to say?*
> **Smithy:** *Yes—f—k off.*
> *Now generally I respect subjects' wishes. . . . But on this occasion, I felt bloodyminded. It had been a frustrating day: a long drive, nothing to show. . . . I persisted.*
> **Me:** *What I'm hoping is that this research will be of some use in. . .*
> **Smithy:** *I'm not f—king interested in what you say your work's about. I told you to f—k off.*
> **Me:** *Look, I tell you what. I won't even ask you any questions. Just talk to me for five minutes—about whatever you like, then I'll f—k off.*
> *It seemed to me that his expression softened. . . . Perhaps, too, I had passed some test of maleness.*
> **Smithy:** *Buy me a drink, and I'll talk to you.*

It turned out that the pub where Levinson took "Smithy" for a pint had a "No Travellers" sign [i.e., Romani people not allowed] on its door, so Levinson's interview also bought Smithy access to an otherwise forbidden place.

Each of the following issues should be considered in a qualitative research project.

Voluntary Participation

The first step in ensuring subjects are participating in a study voluntarily is to provide clear information about the research and the opportunity to decline to participate (King and Horrocks 2010:99). However, even when such information is provided, the ongoing interaction between the researcher and the participant can blur the meaning of voluntary participation:

> A skilled researcher can establish rapport and convince subjects to reveal almost anything, including things the researcher may not want to be responsible for knowing. (Sieber and Tolich 2013:164)

This problem of diminishing attention to the voluntary nature of participation can be reduced when qualitative interviews are conducted in multiple sessions by making clear to participants at each session their opportunity to withdraw consent (King and Horrocks 2010:115).

Maintaining the standard of voluntary participation can present more challenges in participant observation studies, particularly those using an ethnographic approach that immerses the researcher in the social world of the regular participants. Few researchers or institutional review boards (IRBs) are willing to condone covert participation because it offers no way to ensure that participation by the subjects is voluntary. However, interpreting the standard of voluntary participation can be difficult even when the researcher's role is more open. Practically, much field research would be impossible if participant observers were required to request the permission of everyone in a group or setting, no matter how minimal their involvement (Calvey 2017:139). Also, consent is an ongoing process (Levinson 2010; Iphofen and Tolich 2018). As Levinson (2010:197) explains, access to participants is not "some immutable, fixed state" that is resolved through gaining informed consent at the start of a project:

> The implication being that, once negotiated, access ceases to be an issue. Experience soon taught me that . . . what seemed acceptable one day was evidently not the next. Factors here ranged from external events—death in the family, illness, . . . to the mood of respondents on a given day. . . . Sometimes, I suspected people had been told not to speak to me.

Some researchers recommend adherence to a **process consent** standard to adjust the ideal of informed consent to the reality of the evolving nature of social relationships and research plans in qualitative projects. Process consent allows participants to change their decision about participating at any point by requiring the researcher to check with participants at each stage of the project about their willingness to continue in the project (Sieber and Tolich 2013:139; Lahman 2018:78). For example, before the conclusion of his ethnographic study of sidewalk book vendors in New York City, Mitch Duneier (1999) rented a room in which he read to participants the excerpts from his planned book that pertained to them and let them decide whether to give consent at that point.

Subject Well-Being

Every field researcher should consider carefully before beginning a project how to avoid harm to subjects and how to balance the benefits and harms of research (Iphofen and Tolich 2018).

Direct harm to the reputations or feelings of particular individuals is a primary concern. Researchers can avoid such harm, in part, by maintaining the confidentiality of research subjects. Researchers also must try to avoid affecting adversely the course of events while engaged in a setting. For example, concerns were raised about Sudhir Venkatesh's (2008) doctoral research about a gang in a Chicago housing project. Venkatesh had gained the trust of housing residents after being given access to them by a local gang leader and the project's manager. Venkatesh was able to ask residents about their sources of unreported income, such as through babysitting, prostitution, and car repair. Venkatesh then reviewed with the gang leader what he had learned, leading the gang leader to extort more money from residents who had sources of income about which he had been unaware and leading Venkatesh to be shunned by the residents who had shared their income secrets with him (Sieber and Tolich 2013:88–90). Although such consequences could follow from any research, even from any public discourse, they are a particular concern for qualitative researchers who become during their research an accepted part of a group or community.

These problems are less likely in intensive interviewing and focus groups than ethnography, but researchers using these methods should try to identify negative feelings both after interviews and after reports are released and then help distressed subjects cope with their feelings through debriefing or referrals for professional help.

Identity Disclosure

We already have considered the problems of identity disclosure, particularly in the case of covert participation. Current ethical standards require informed consent of research subjects, and most would argue this standard cannot be met in any meaningful way if researchers do not disclose fully their identity. But how much disclosure about the study is necessary, and how hard should researchers try to make sure their research purposes are understood? For example, in his ethnography of the Dark Web Social Network (DWSN), Gehl's (2016) participants, in keeping with the privacy and anonymity-focused culture of the DSWN, refused to be informed of Gehl's full identity. Also, less-educated subjects may not readily comprehend what a researcher is or be able to weigh the possible consequences of the research for themselves. There are several instances in which the level and frequency of identity disclosure required is an ethical challenge.

Consider this: In her study of female gamblers, Jun Li (2008) felt she could not ethically maintain her covert status when older women gamblers made concerted efforts to dissuade her from getting addicted. However, when Li disclosed her identity as a researcher, they stopped talking to her:

> Once female gamblers were made known of my research role, they started to view me differently, treating me as a suspicious outsider who should not be entrusted because I did not share their experiences. (P. 107)

Digital ethnographers also have to be concerned with identity disclosure by participants in online communities who can easily misrepresent their own identities. Men may masquerade as women and children as adults, or vice versa, and yet the research should not proceed unless an effort is made to obtain informed and voluntary consent. If they are concerned about participation by individuals who do not meet criteria for a study, including minimal age requirements, digital ethnographers can at least attempt to identify those whose identity is not credible through inconsistencies in their postings (Kozinets 2010:151–154). Internet-based research can also violate the principles of voluntary participation and identity disclosure when researchers participate in discussions and record and analyze text but do not identify themselves as researchers (Jesnadum 2000). Digital ethnographers minimize the risks caused by these uncertainties

by making their own identities known, stating clearly their expectations for participation, and providing an explicit informed consent letter that is available as discussion participants come and go (Denzin and Lincoln 2008:274). Janet Salmons (2016:11) recommends communicating research plans to the website host and posting an introduction to the study written in plain language.

Confidentiality

Field researchers normally use fictitious names for the characters in their reports, but doing so does not always guarantee confidentiality to their research subjects. Individuals in the setting studied may be able to identify those whose actions are described and thus may become privy to some knowledge about their colleagues or neighbors that had formerly been kept from them. Therefore, researchers should make every effort to expunge possible identifying material from published information and to alter unimportant aspects of a description when necessary to prevent identity disclosure. In any case, no field research project should begin if some participants clearly will suffer serious harm by being identified in project publications. On the other hand, the American Sociological Association does not require informed consent from persons in public places who are being observed as a condition of research, although local IRBs may differ (Sieber and Tolich 2013:137–138).

Focus groups create a particular challenge because the researcher cannot guarantee that participants will not disclose information that others would like to be treated as confidential. Cultures differ in their openness to dissent and their sensitivity to public discussion of personal issues, so cross-cultural and cross-national focus groups exacerbate the difficulty of establishing appropriate standards for confidentiality (Smithson 2008:365–366). This risk can be reduced at the start of the focus group by reviewing a list of "dos and don'ts" with participants. Nonetheless, focus group methodology simply should not be used for very personally sensitive topics (Smithson 2008:360–361).

Although a digital ethnographer should maintain the same privacy standards as other qualitative researchers do, the retention of information on the web makes it difficult to protect confidentiality with traditional mechanisms. Anyone reading a distinctive quote based on online text can use a search engine to try to locate the original text. Distinctive online names intended to protect anonymity still identify individuals who may be the target of hurtful commentary (Kozinets 2010:143–145). Some users may think of their postings to some online community site as private (Markham 2008:274). For these reasons, digital ethnographers who study sensitive issues should go to great lengths to disguise the identity of the community they have studied as well as of its participants who are quoted.

Appropriate Boundaries

Maintaining appropriate boundaries between the researcher and research participants is a uniquely important issue in qualitative research projects that creates challenges for identity disclosure, subject well-being, and voluntary participation. You probably are familiar with this issue in the context of guidelines for ethical professional practice that seek to prevent abuse resulting from the inherent power disparities in relations between clinicians and patients and teachers and students. This is a special issue in qualitative research that involves reducing the distinction between the researcher and the research subject (or research participant). Qualitative researchers may seek to build rapport with those they plan to interview by expressing an interest in their concerns and conveying empathy for their situation.

> With deeper rapport, interviewees become more likely to explore their more intimate experiences and emotions. Yet they also become more likely to discover and disclose experiences and feelings which, upon reflection, they would have preferred to keep private from others . . . , or not to acknowledge even to themselves. (Duncombe and Jessop 2002:112)

Are researchers just "faking friendship" for the purpose of the research, as Jean Duncombe and Julie Jessop (2002) posed the dilemma in a book chapter titled "'Doing Rapport' and the Ethics of 'Faking Friendship'"? The long-term relationships that can develop in participant observation studies can make it seem natural for researchers to offer tangible help to research participants, such as helping take a child to school or lending funds. These involvements can in turn make it difficult to avoid becoming an advocate for the research participants rather than a sympathetic observer. Duneier (1999) explains how he grappled with this issue and the way he resolved it in his study of impoverished sidewalk book vendors in New York City:

> Could I show my deep appreciation for their struggles and gain their appreciation for my purposes as a sociologist without paying for some simulacrum of it? How could I communicate my purposes as a researcher without dollar bills and small change in my hand? . . . I knew that my salary (while not very high) was quite high compared to the going rate on the sidewalk. . . . But with time I did learn to say no, and to communicate the anguish I felt in giving such an answer.
>
> The question of how to avoid intervening when one cannot or should not do so is different from the question of whether and how to help when one can and should. At times I was asked to do things as simple as telling what I knew about the law, serving as a reference for a person on the sidewalk as he or she dealt with a landlord or potential landlord, helping someone with rent when he was about to be evicted, and on one occasion finding and paying for a lawyer. In these situations, I did everything I could to be helpful, but I never gave advice, opinions, or help beyond what was asked for. (p. 355)

Researcher Safety

Research "in the field" about disasters or simply unfamiliar neighborhoods or nations should not begin until any potential risks to researcher safety have been evaluated. As Virginia Dickson-Swift et al. (2008) note, risks in qualitative research can be emotional as well as physical:

> Look after yourself—have someone who you can debrief with, in fact have two people who you can debrief with. . . . Think there is always a risk in this research. You have got to remember that sometimes we go into spaces in people's lives that others have not been and that has the potential to be risky, both physically if the environment is not a safe one, or emotionally if the research affects you in some way. (pp. 137–138)

Qualitative methods may provide the only opportunity to learn about organized crime in Russian ports (Belousov et al. 2007) or street crime in the Dominican Republic (Gill 2004), but they should not be used if the physical risks to the researchers are unacceptably high. Safety needs to be considered at the time of designing the research, not as an afterthought on arriving at the research site. As Gill (2004) learned, such advance planning can require more investigation than just reading the local newspapers: "Due to the community's marginality, most crimes, including murders, were never reported in newspapers, making it impossible to have known the

insecurity of the field site ahead of time" (p. 2). But being realistic about evaluating risk does not mean simply accepting misleading assumptions about unfamiliar situations or communities.

These ethical issues cannot be evaluated independently. The final decision to proceed must be made after weighing the relative benefits and risks to participants and considering how throughout the project voluntary participation can reasonably be ensured, researcher identity can be disclosed to newcomers, researcher-induced harm can be largely prevented, confidentiality maintained, and risks and benefits clearly and honestly evaluated. Few qualitative research projects will be barred by consideration of these ethical issues, but almost all projects require careful attention to them. The more important concern for researchers is to identify the ethically troublesome aspects of their proposed research and resolve them before the project begins and to act on new ethical issues as they come up during the project.

CONCLUSIONS

Qualitative research allows the careful investigator to obtain a richer and more intimate view of the social world than is possible with more structured methods. It is not hard to understand why so many qualitative studies have become classics in the social science literature. And the emphases in qualitative research on inductive reasoning and incremental understanding help stimulate and inform other research approaches. Exploratory research to chart the dimensions of previously unstudied social settings and intensive investigations of the subjective meanings that motivate individual action are particularly well served by the techniques of participant observation, intensive interviewing, and focus groups.

The very characteristics that make qualitative research techniques so appealing restrict their use to a limited set of research problems. It is not possible to draw representative samples for study using participant observation, and, for this reason, the generalizability of any particular field study's results cannot really be known. Only the accumulation of findings from numerous qualitative studies permits confident generalization, but here again, the time and effort required to collect and analyze the data make it unlikely that many field research studies will be replicated.

Even if qualitative researchers made more of an effort to replicate key studies, their notion of developing and grounding explanations inductively in the observations made in a particular setting would hamper the comparison of findings. Measurement reliability is thereby hindered, as are systematic tests for the validity of key indicators and formal tests for causal connections.

In the final analysis, qualitative research involves a mode of thinking and investigating different from that used in experimental and survey research. Qualitative research is inductive and idiographic, whereas experiments and surveys tend to be conducted in a deductive, quantitative, and nomothetic framework. Both approaches can help social scientists learn about the social world; the proficient researcher must be ready to use either. Qualitative data are often supplemented with counts of characteristics or activities. And as you have already seen, quantitative data are often enriched with written comments and observations, and focus groups have become a common tool of survey researchers seeking to develop their questionnaires. Thus, the distinction between qualitative and quantitative research techniques is not always clear-cut, and combining methods is often a good idea. I'll return to this in Chapter 12, on mixed methods.

KEY TERMS

Adaptive research design (p. 355)
Case study (p. 355)
Complete (or covert) participant
Complete (or overt) observer
Constructivism
Covert observer (p. 358)
Digital ethnography (p. 357)
Ethnography (p. 356)
Field notes (p. 367)
Field research (p. 351)
Field researcher
Focus groups
Gatekeeper (p. 363)
Grand tour question
Hermeneutic circle (p. 354)
Intensive (in-depth) interviewing (p. 369)
Jottings (p. 367)
Key informant
Participant observation (p. 358)
Participant observer (p. 358)
Process consent (p. 377)
Reactive effects (p. 360)
Reflexivity (p. 353)
Saturation point (p. 371)
Theoretical sampling (p. 366)
Thick description (p. 355)

HIGHLIGHTS

LO 10.1 Recall the features of qualitative research that most distinguish it from experimental and survey research designs.

- Qualitative research is the collection primarily of qualitative rather than quantitative data.
- Qualitative research focuses on previously unstudied processes and unanticipated phenomena.
- Qualitative research pursues exploratory research questions, with a commitment to inductive reasoning.
- Qualitative research is sensitive to the subjective role of the researcher.
- Qualitative research is is oriented to social context, to the interconnections between social phenomena rather than to their discrete features.
- Qualitative research has a focus on human subjectivity (on the meanings that participants attach to events and that people give to their lives).
- Qualitative research uses idiographic rather than nomothetic causal explanations.
- Some qualitative researchers accept a constructivist philosophy.
- Qualitative research often includes adaptive research design, in which the design develops as the research progresses.

LO 10.2 Compare and contrast the styles of qualitative research.

- A case study is a setting or group the analyst treats as an integrated social unit that must be studied holistically and with concern for its particularity. Central to much qualitative case study research is the goal of creating a thick description of the setting studied.
- Traditional ethnography is the study of a culture or cultures that some group of people shares, using participant observation over an extended period.
- Digital ethnographers use ethnographic techniques to study online communities.
- Hybrid ethnographers use both traditional and digital ethnographic methods to capture both the in-person and online dimensions of a social setting.

LO 10.3 Identify important considerations when conducting participant observation.

- *Choosing a role*. The first concern of every participant observer is to decide what balance to strike between observing and participating and whether to reveal one's role as a researcher. Researchers using participant observation can adopt one of several different roles in a setting: covert observer, complete observer, complete participant, or participant observer.
- *Entering the field*. One must do some background work on the social norms (e.g., social boundaries within a setting, how participants tend to dress, and common activities among participants) of a setting before entering the field in order to make a good impression on participants.
- *Developing and maintaining relationships*. One must be careful to manage their relationships in the research setting so they can continue to observe and interview diverse members of the social setting throughout the long period typical of participant observation.
- *Sampling people and events.* Most qualitative researchers limit their focus to just one or a few sites, so they can focus all their attention on the social dynamics of those settings. The researcher must be reasonably confident about gaining access and that the site can provide relevant information. The sample must be appropriate and adequate for the study, even if it is not representative.
- *Taking notes.* Written notes are the primary means of recording participant observation data. Many field researchers jot down partial notes while observing and then retreat to their computers to write up more complete notes daily. Careful note-taking yields a big payoff. On page after page, field notes will suggest new concepts, causal connections, and theoretical propositions.
- *Managing the personal dimensions*. Because field researchers become a part of the social situation they are studying, they cannot help but be affected on a personal, emotional level. At the same time, those being studied react to researchers not just as researchers but as personal acquaintances—often as friends, sometimes as personal rivals. Managing and learning from this personal side of field research is an important part of any project.

LO 10.4 Compare the process of intensive interviewing to the process of interviewing in survey research.

- Intensive or depth interviewing is a qualitative method of finding out about people's experiences, thoughts, and feelings.
- Unlike the more structured interviewing that may be used in survey research, intensive or depth interviewing relies on open-ended questions.
- Rather than asking standard questions in a fixed order, intensive interviewers may allow the specific content and order of questions to vary from one interviewee to another.
- Rather than presenting fixed responses that presume awareness of the range of answers that respondents might give, intensive interviewers expect respondents to answer questions in their own words.
- Intensive interviews are often much longer than standardized interviews, sometimes as long as 15 hours, conducted in several different sessions.
- Intensive interviewing engages researchers more actively with respondents than standard survey research does.

LO 10.5 Assess how conversations within focus groups are similar to and different from ordinary, non-research conversations.

- Focus groups appear to be similar to ordinary, non-research conversations because of the following:
  - They are informal, engaging, and friendly discussions among a group of people.
  - A moderator may not be particularly directive, and there can be freedom of conversation.
- Focus groups are different from ordinary, non-research conversations because of the following:
  - People are selected to participate in focus groups based on specific characteristics.
  - The moderator usually begins a discussion by generating interest in the topic, creating the expectation that all will participate and making it clear they do not favor any particular perspective or participant.
  - The moderator often strategically begins with easy-to-answer general factual questions and then, about one quarter to halfway through the allotted time, shifts to key questions on specific issues.
  - Moderators are specifically trained to handle group conflicts and keep the group conversation focused on the topic of interest.

LO 10.6 Argue whether it makes sense to be concerned with generalizability in qualitative research.

- Although nonprobability sampling methods are often used in qualitative research, many researchers argue that generalizability should not be ignored.
- Janet Ward Schofield (2002) outlines two ways to increase generalizability in qualitative research: studying the typical and performing multisite studies.
- Giampietro Gobo (2008) describes how generalizability in qualitative research can be improved by selecting an atypical, or deviant, case for in-depth study. Investigating social processes in a situation that differs from the norm will improve understanding of how social processes work in typical situations: "the exception proves the rule."
- Still, some qualitative researchers question the value of generalizability; the argument is that understanding the particulars of a situation in depth is an important goal in itself.

LO 10.7 Identify the major ethical challenges faced by qualitative researchers.

- *Voluntary participation*. The first step in ensuring subjects are participating in a study voluntarily is to provide clear information about the research and the opportunity to decline to participate. Participants should be reminded of the voluntary nature of research participation on a regular basis during ongoing interactions (e.g., process consent).
- *Subject well-being*. Direct harm to the reputations or feelings of particular individuals is a primary concern. Researchers can avoid such harm, in part, by maintaining the confidentiality of research subjects. Researchers also must try to avoid affecting adversely the course of events while engaged in a setting.
- *Identity disclosure*. Current ethical standards require informed consent of research subjects, and most would argue this standard cannot be met in any meaningful

way if researchers do not disclose fully their identity. Internet-based research can also violate identity disclosure when researchers participate in discussions and record and analyze text but do not identify themselves as researchers. Also, digital ethnographers have to be concerned with identity disclosure by participants in online communities who can easily misrepresent their own identities.

- *Confidentiality.* It is difficult to be able to guarantee complete confidentiality to research subjects (e.g., individuals in the setting studied may be able to identify those whose actions are described; focus group members may disclose information that other members would like to be treated as confidential). Researchers should make every effort to expunge possible identifying material from published information and to alter unimportant aspects of a description when necessary to prevent identity disclosure.
- *Appropriate boundaries.* The long-term relationships that can develop in participant observation studies can make it seem natural for researchers to offer tangible help to research participants, such as helping take a child to school or lending funds. These involvements can in turn make it difficult to avoid becoming an advocate for the research participants rather than a sympathetic observer.
- *Researcher safety.* Safety needs to be considered at the time of designing the research, not as an afterthought on arriving at the research site. Researchers must weigh whether the risks are unacceptably high in a qualitative research setting. Risks to safety in qualitative research can be emotional as well as physical.

DISCUSSION QUESTIONS

1. What aspects of the social world would be difficult for participant observers to study without being covert? Are there any situations that would require the use of covert observation to gain access? What might you do as a participant observer to lessen access problems while still acknowledging your role as a researcher?

2. Review the experiments and surveys described in previous chapters. Pick one and propose a field research design that would focus on the same research question but with participant observation techniques in a local setting. Propose the role you would play in the setting, along the participant observation continuum, and explain why you would favor this role. Describe the stages of your field research study, including your plans for entering the field, developing and maintaining relationships, sampling, and recording and analyzing data. Then, discuss what you would expect your study to add to the findings resulting from the study described in the book.

3. Intensive interviews are the core of many qualitative research designs. How do they differ from the structured survey procedures you studied in Chapter 8? What are their advantages and disadvantages over standardized interviewing? How does intensive interviewing differ from the qualitative method of participant observation? What are the advantages and disadvantages of these two methods?

4. Research on disasters poses several methodological challenges. In what ways are qualitative methods suited to disaster research? What qualitative methods would you have emphasized if you had been able to design research in New Orleans in the immediate aftermath of Hurricane Ida? What unique challenges would you have confronted because of the nature of the disaster?

SPSS EXERCISES

The cross-tabulations you examined in Chapter 8's SPSS exercises highlighted the strength of the association between attitudes related to race and support for capital punishment. In this chapter, you will explore related issues. As you carry out this analysis, consider what additional information you might be able to obtain about these relationships with qualitative interviews of some of the respondents.

1. Examine the association between race and support for capital punishment. From the menu, click:
   Analyze/Descriptive Statistics/Crosstabs
   In the Crosstabs window, set
   Rows: **CAPPUN_2**
   Columns: **RACE2020**
   Cells: **column percents**

2. What is the association between race and support for capital punishment? How would you explain that association?

3. Now consider what might lead to variation in support for capital punishment between whites and Blacks. Consider gun ownership (OWNGUN_2), religious beliefs (FUND2020), attitudes about race (GOVAIDBLK3), education (EDUC5), and political party identification (PARTYID4).

4. Generate crosstabs for the association of support for capital punishment with each of these variables, separately for Blacks and whites. Follow the same procedures you used in Step 1, substituting the variables mentioned in Step 3 for RACE2020 in Step 1. However, you must repeat the crosstab request for Blacks and whites. To do this, before you choose Analyze, select white respondents only. From the menu above the Data Editor window, select Data, then Select Cases. Then from the Select Cases window, select "if condition is satisfied" and create this expression:

   If... **RACE2020 = 1**

   After you have generated the crosstabs, go back, and repeat the data selection procedures, ending with RACE2020 = 2. When finished with the exercises, be sure to go back to Select Cases and select All Cases.

5. Are the bases of support for capital punishment similar among Blacks and whites? Discuss your findings.

6. Propose a focus group to explore these issues further. Identify the setting and sample for the study and describe how you would carry out your focus group.

©iStockphoto.com/Aldo Murillo

11 QUALITATIVE DATA ANALYSIS

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

11.1 Explain the meaning of an emic focus and of an etic focus in qualitative research

11.2 List the five different techniques that are shared by most approaches to qualitative data analysis

11.3 Defend the value of using more than one method of analyzing qualitative data in a project

11.4 Identify changes in the social world that have led to the growth of visual sociology

11.5 Argue for the strengths of a systematic observation technique

11.6 Appraise the strengths of participatory action research

11.7 Recommend the ways in which computer-assisted qualitative data analysis can facilitate research

11.8 Review the following ethical issues that should be given special attention in qualitative data analysis: (1) privacy, confidentiality, and anonymity; (2) research integrity and quality; (3) ownership of data and conclusions; and (4) use and misuse of results

My dad used to work construction and that was his only employment, and so he would get home at 7 or 8 at night and I can't imagine going home after a long day of laborious work and still have the energy to play with your kids. (laughs) . . . He would play with us when he got home. Not all the time but he would. He was very unselfish and he joined the weekends when he was at home, he would, you know, going out to have dinner and having four kids it's expensive, you know, it's expensive. (Coleman-Minahan and Samari 2020:568)

Among the many hazards in migrating to the United States from Mexico are strains in parent-child relationships. Between March 2013 and April 2014, Kate Coleman-Minahan (Coleman-Minahan and Samari 2020) set out to investigate these relationships by conducting life-history interviews with 21 Mexican-origin young women in immigrant families in the Denver area. In their subsequent qualitative data analysis, she and coauthor Goleen Samari identified four types of father–daughter relationships. They use this quote to illustrate the type of father–daughter relationship they classify as good.

In this chapter, I present the features that most qualitative data analyses share. You will quickly learn that there is no one way to analyze qualitative, textual, or visual data. To quote Michael Quinn Patton (2002),

> Qualitative analysis transforms data into findings. No formula exists for that transformation. Guidance, yes. But no recipe. Direction can and will be offered, but the final destination remains unique for each inquirer, known only when—and if—arrived at. (p. 432)

I will then distinguish different types of qualitative data analysis and introduce computer programs for qualitative data analysis—increasingly popular tools for qualitative data analysis that are blurring the distinctions between quantitative and qualitative approaches to textual analysis.

FEATURES OF QUALITATIVE DATA ANALYSIS

The distinctive features of qualitative data collection methods that you studied in Chapter 10 are also reflected in the methods used to analyze those data. The focus on text—on qualitative data rather than on numbers—is the most important feature of qualitative analysis. The "text" that qualitative researchers analyze is most often transcripts of interviews or notes from participant observation sessions, but *text* can also refer to pictures or other images the researcher examines.

What can the qualitative data analyst learn from a text? Qualitative analysts may have two different goals. Some qualitative researchers view analysis of a text as a way to understand what participants "really" thought, felt, or did in a given situation or at a point in time. The text becomes a way to get "behind the numbers" that are recorded in a quantitative analysis to see the richness of real social experience. Others have adopted a hermeneutic perspective on texts—that is, a perspective that views a text as an interpretation that can never be judged true or false. The text is only one possible interpretation among many (Patton 2002:114).

From a hermeneutic perspective, then, the meaning of a text is negotiated among a community of interpreters, and to the extent that some agreement is reached about meaning at a particular time and place, that meaning is based on consensual community validation. A hermeneutic researcher is thus constructing a "reality" with their interpretations of a text provided by the subjects of research; other researchers, with different backgrounds, could come to markedly different conclusions.

Qualitative and quantitative data analyses tend to differ in the priority given to the preexisting views of the researcher and to those of the subjects of the research. Qualitative data analysts seek to describe their textual data in ways that capture the setting or people who produced this text on their own terms rather than in terms of predefined measures and hypotheses. What this means is that qualitative data analysis is more likely to be inductive—the analyst identifies important categories in the data, as well as patterns and relationships, through a process of discovery. There are often no predefined measures or hypotheses. Anthropologists term this an **emic focus**, which means representing the setting in terms of the participants and their viewpoints—an "insider" perspective based on the researcher's immersion in the setting. This focus contrasts with the **etic focus** of most quantitative research, in which the researcher seeks to remove themself from the setting and represents the participants in the researcher's terms and in relation to a research question based on the preexisting literature (Salmons 2016:102).

Good qualitative data analyses also are distinguished by their focus on the interrelated aspects of the setting, group, or person under investigation—the case—rather than breaking the whole into separate parts. The whole is always understood to be greater than the sum of its parts, and so the social context of events, thoughts, and actions becomes essential for interpretation. Within this framework, it doesn't really make sense to focus on two variables out of an interacting set of influences and test the relationship between just those two.

Qualitative data analysis is an iterative and reflexive process that begins as data are being collected rather than after data collection has ceased (Stake 1995). Next to their field notes or interview transcripts, the qualitative analyst jots down ideas about the meaning of the text and how it might relate to other issues. This process of reading through the data and interpreting them continues throughout the project. The analyst adjusts the data collection process itself when it begins to appear that additional concepts need to be investigated or new relationships explored. This process is termed **progressive focusing** (Parlett and Hamilton 1976).

> We emphasize placing an interpreter in the field to observe the workings of the case, one who records objectively what is happening but simultaneously examines its meaning

> and redirects observation to refine or substantiate those meanings. Initial research questions may be modified or even replaced in mid-study by the case researcher. The aim is to thoroughly understand [the case]. If early questions are not working, if new issues become apparent, the design is changed. (Stake 1995:9)

Elijah Anderson (2003) describes progressive focusing in a memoir about his study of Jelly's Bar:

> Throughout the study, I also wrote conceptual memos to myself to help sort out my findings. Usually no more than a page long, they represented theoretical insights that emerged from my engagement with the data in my field notes. As I gained tenable hypotheses and propositions, I began to listen and observe selectively, focusing on those events that I thought might bring me alive to my research interests and concerns. This method of dealing with the information I was receiving amounted to a kind of a dialogue with the data, sifting out ideas, weighing new notions against the reality with which I was faced there on the streets and back at my desk. (pp. 235–236)

Carrying out this process successfully is more likely if the analyst reviews a few basic guidelines when they start the process of analyzing qualitative data (Miller and Crabtree 1999b:142–143):

- Know yourself, your biases, and preconceptions.
- Know your question.
- Seek creative abundance. Consult others and keep looking for alternative interpretations.
- Be flexible.
- Exhaust the data. Try to account for all the data in the texts, then publicly acknowledge the unexplained and remember the next principle.
- Celebrate anomalies. They are the windows to insight.
- Get critical feedback. The solo analyst is a great danger to self and others.
- Be explicit. Share the details with yourself, your team members, and your audiences.

Note particularly that the qualitative analyst is not expected to cross the finish line alone. "Consult others," "publicly acknowledge the unexplained," "get critical feedback," and "share the details" all emphasize the engagement of the qualitative data analyst in a community of practice—other researchers who share similar interests and perspectives—from whom critical advice is sought. Some suggest using a "critical friend" to ensure the analyst confronts critical issues and recognizes their own biases (Blanco and Rossman 2022).

Also important is the element of flexibility. The analyst who keeps looking for alternative interpretations and seeks new "insight" is able to redesign the analytic process and even seek more interviews or new observations as new interpretations or questions emerge (Blanco and Rossman 2022:19–20).

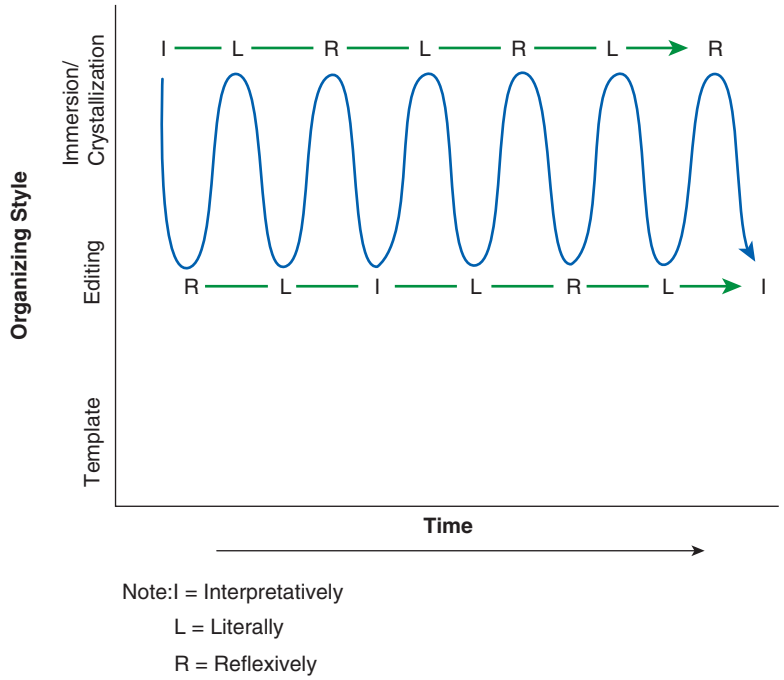
Qualitative Data Analysis as an Art

If you find yourself longing for the certainty of predefined measures and deductively derived hypotheses, you are beginning to understand the difference between setting out to analyze data

quantitatively and planning to do so with a qualitative approach in mind. Or maybe you are now appreciating better the contrast between the positivist and constructivist research philosophies I summarized in Chapter 1. When it comes right down to it, the process of qualitative data analysis is even described by some as involving as much "art" as science—as a "dance," in the words of William Miller and Benjamin Crabtree (1999b) (see Figure 11.1):

> Interpretation is a complex and dynamic craft, with as much creative artistry as technical exactitude, and it requires an abundance of patient plodding, fortitude, and discipline. There are many changing rhythms; multiple steps; moments of jubilation, revelation, and exasperation. . . . The dance of interpretation is a dance for two, but those two are often multiple and frequently changing, and there is always an audience, even if it is not always visible. Two dancers are the interpreters and the texts. (pp. 138–139)

FIGURE 11.1 ■ Dance of Qualitative Analysis



Source: Miller and Crabtree (1999b:139, Figure 7.1, based on Addison 1999). Reprinted with permission from SAGE Publications, Inc.

Miller and Crabtree (1999b) identify three different modes of reading the text within the dance of qualitative data analysis:

1. When the researcher reads the text literally, they are focused on its literal content and form, so the text "leads" the dance.
2. When the researcher reads the text reflexively, they focus on how their own orientation shapes their interpretations and focus. Now, the researcher leads the dance.
3. When the researcher reads the text interpretively, they try to construct their own interpretation of what the text means.

Sherry Turkle's (2011) book *Alone Together: Why We Expect More From Technology and Less From Each Other* provides many examples of this analytic dance, although of course in the published book we are no longer able to see that dance in terms of Turkle's original notes. She often describes what she observed in classrooms. Here's an example of such a *literal* focus, reflecting her experience in MIT's Media Lab at the start of the mobile computing revolution:

> In the summer of 1996, I met with seven young researchers at the MIT Media Lab who carried computers and radio transmitters in their backpacks and keyboards in their pockets. . . . They called themselves "cyborgs" and were always wirelessly connected to the Internet, always online, free from desks and cables. (Turkle 2011:151)

Such literal reports are interspersed with *interpretive* comments about the meaning of her observations:

> The cyborgs were a new kind of nomad, wandering in and out of the physical real. . . . The multiplicity of worlds before them set them apart; they could be with you, but they were always somewhere else as well. (Turkle 2011:152)

And several times in each chapter, Turkle (2011) makes *reflexive* comments on her own reactions:

> I don't like the feeling of always being on call. But now, with a daughter studying abroad who expects to reach me when she wants to reach me, I am grateful to be tethered to her through the Net. . . . Even these small things allow me to identify with the cyborgs' claims of an enhanced experience. Tethered to the Internet, the cyborgs felt like more than they could be without it. Like most people, I experience a pint-sized version of such pleasures. (p. 153)

In this artful way, the qualitative data analyst reports on their notes from observing or interviewing, interprets those notes, and considers how they react to the notes. These processes emerge from reading the notes and continue while they are editing the notes and deciding how to organize them, in an ongoing cycle.

Qualitative Compared With Quantitative Data Analysis

With this process in mind, let's review the many ways in which qualitative data analysis differs from quantitative analysis (Denzin and Lincoln 2000:8–10; Patton 2002:13–14). Each difference reflects the qualitative data analyst's orientation to in-depth, comprehensive understanding in which the analyst is an active participant compared with the quantitative data analyst's role as a dispassionate investigator of specific relations between discrete variables:

- A focus on meanings rather than on quantifiable phenomena
- Collection of many data on a few cases rather than few data on many cases
- Study in depth and detail, without predetermined categories or directions, rather than emphasis on analyses and categories determined in advance
- Conception of the researcher as an "instrument" rather than as the designer of objective instruments to measure particular variables
- Sensitivity to context rather than seeking universal generalizations

- Attention to the impact of the researcher's and others' values on the course of the analysis rather than presuming the possibility of value-free inquiry
- A goal of rich descriptions of the world rather than measurement of specific variables

The focus of qualitative data analysis on meaning and in-depth study also makes it a valuable supplement to analyses of quantitative data. Qualitative data like that collected in Coleman-Minahan and Samari's (2020) qualitative interviews of young women in Mexican-origin families in and around Denver, Colorado, provided much information about the quality of father–daughter relationships. This qualitative information helped to explain the association Coleman-Minahan and Samari's (2020) found between father–daughter relationship quality and reduced risk of daughters' early sexual initiation in an analysis of quantitative survey data.

You'll also want to keep in mind features of qualitative data analysis that are shared with those of quantitative data analysis. Both qualitative and quantitative data analysis can involve making distinctions about textual data. Textual data can be transposed to quantitative data through a process of categorization and counting (see Chapter 14 on content analysis). Some qualitative analysts also share with quantitative researchers a positivist goal of describing better the world as it "really" is, although others have adopted a postmodern goal of trying to understand how different people see and make sense of the world, without believing there is any "correct" description.

TECHNIQUES OF QUALITATIVE DATA ANALYSIS

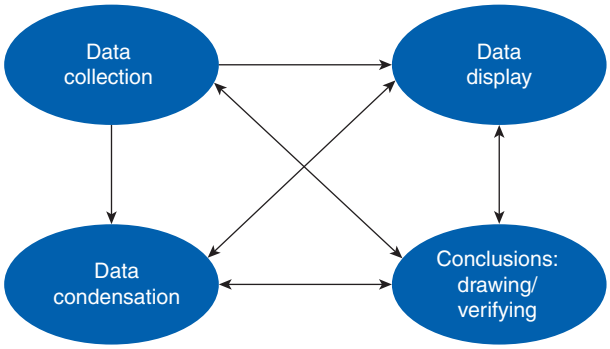
Five different techniques are shared by most approaches to qualitative data analysis:

1. Documentation of the data and the process of data collection
2. Organization, categorization, and condensation of the data into concepts
3. Examination and display of relationships between concepts
4. Corroboration and legitimization of conclusions, by evaluating alternative explanations, disconfirming evidence, and searching for negative cases
5. Reflection on the researcher's role

Some researchers suggest different steps or include additional steps, such as developing propositions that reflect the relationships found and making connections with extant theories (see Miles, Huberman, and Soldaña 2014, chap. 1). Figure 11.2 highlights the key techniques and emphasizes the reciprocal relations between them. In qualitative data analysis, condensation of data into concepts may lead to some conclusions and to a particular form of display of relationships between concepts, but the conclusions may then lead to changes in conceptualization and display, in an iterative process.

The analysis of qualitative research notes begins in the field, at the time of observation, interviewing, or both, as the researcher jots down problems and concepts that appear likely to help in understanding the situation being observed or described. Researchers should make frequent notes—"observer's comments" (OCs)—in the margins to identify important statements and to propose ways of coding the data: "husband–wife conflict," perhaps, or "tension-reduction

FIGURE 11.2 ■ Components of Data Analysis: Interactive Model



Source: Miles, Huberman, and Saldaña (2014, chap. 1). Reprinted with permission from SAGE Publications, Inc.

strategy" (Saldana and Omasta 2022:38). Simply reading the notes or transcripts after an observation session or interview is also an important step in the analytic process.

An interim stage may consist of listing the concepts reflected in the notes and diagramming the relationships between concepts (Maxwell 2005:97–99). In large projects, weekly team meetings are an important part of this process. Susan Miller (1999) described this process in her study of neighborhood police officers (NPOs). Her research team met both to go over their field notes and to resolve points of confusion, as well as to dialogue with other skilled researchers who helped identify emerging concepts:

> The fieldwork team met weekly to talk about situations that were unclear and to troubleshoot any problems. We also made use of peer-debriefing techniques. Here, multiple colleagues, who were familiar with qualitative data analysis but not involved in our research, participated in preliminary analysis of our findings. (p. 233)

This process continues throughout the project and should assist in refining concepts during the report-writing phase, long after data collection has ceased. Let's examine each of the stages of qualitative research in more detail.

Documentation

The data for a qualitative study most often are text transcribed from audio recordings or notes jotted down in the field (perhaps supplemented with pictures)—from which the original comments, observations, and feelings are reconstructed. In a case study, some qualitative researchers suggest that at least 20 hours of participant observation is required to collect sufficient data for understanding the setting. Ethnographers often expect between a year to a year and a half of daily fieldwork to become sufficiently familiar with a culture (Saldana and Omasta 2022:41). Just a single 1-hour interview can generate 20–25 pages of single-spaced text (Kvale 1996:169). What to do with all this material? Many field research projects have slowed to a halt because a novice researcher becomes overwhelmed by the quantity of information that has been collected. Analysis is less daunting, however, if the researcher maintains a disciplined transcription schedule.

> Usually, I wrote these notes immediately after spending time in the setting or the next day. Through the exercise of writing up my field notes, with attention to "who" the

> speakers and actors were, I became aware of the nature of certain social relationships and their positional arrangements within the peer group. (Anderson 2003:235)

You can see the analysis already emerging from this simple process of taking notes. The transcription process itself can now be partially automated and considerably aided with voice recognition software. Many useful programs provide sufficient accuracy for creating a viable first draft. Online interviews can be transcribed as they happen in Zoom © and other virtual meeting services. Although careful inspection of results is a must, the reduction in transcription time can significantly lessen project costs (di Gregorio 2022).

The first formal analytic step is documentation. The various contacts, interviews, written documents, and whatever it is that preserves a record of what happened all need to be saved and listed. Documentation is critical to qualitative research for several reasons: It is essential for keeping track of what will be a rapidly growing volume of notes, audio—and perhaps video recordings—and documents; it provides a way of developing and outlining the analytic process; and it encourages ongoing conceptualizing and strategizing about the text. Days and times of interviews or observations, locations, and content summaries should be included (Saldana and Omasta 2022:41).

Matthew Miles and A. Michael Huberman (1994:53) provide a good example of a contact summary form that was used to keep track of observational sessions in a qualitative study of a new school curriculum (see Figure 11.3).

Coding, Categorization, and Condensation

Identifying and refining important concepts so they can be organized and categorized is a key part of the iterative process of qualitative research. Conceptual identification can begin with a process of coding, in which a word or short phrase is used to represent a portion of text, an image, or an observation the researcher considers relevant to their research question (Richards 2022:156; Saldana 2021:5). A first step in such coding may be "in vivo coding," which involves identifying words or phrases that stand out in the text as conveying important meaning. The following example illustrates both two researcher-created codes and one in vivo code from a transcript of an interview with a high school student who is describing his favorite teacher.

Codes can also identify processes involving participant action, values that are important to people, and emotions expressed (Saldana and Omasta 2022:112–116).

Sometimes, conceptual organization begins with a simple observation that is interpreted directly, "pulled apart," and then put back together more meaningfully. Robert Stake (1995) provides an example:

> When Adam ran a push broom into the feet of the children nearby, I jumped to conclusions about his interactions with other children: aggressive, teasing, arresting. Of course, just a few minutes earlier I had seen him block the children climbing the steps in a similar moment of smiling bombast. So I was aggregating, and testing my unrealized hypotheses about what kind of kid he was, not postponing my interpreting. . . . My disposition was to keep my eyes on him. (p. 74)

The focus in this conceptualization "on the fly" is to provide a detailed description of what was observed and a sense of why that was important.

In the qualitative analysis process, emerging analytic insights are tested against new observations, the initial statement of problems and concepts is refined, the researcher then collects more data, interacts with the data again, and the process continues. In this way, the analytic process can cycle through both inductive and deductive steps, as the analyst develops ideas from the

FIGURE 11.3 ■ Example of a Contact Summary Form

Site: Tindale

Contact type: _____ Contact date: 11/28-29/79

Visit _____ X_____ Today's date: 12/28/79

Phone _____ Written by: BLT

(with whom)

1. What were the main issues or themes that struck you in this contact?

   Interplay between highly prescriptive, "teacher-proof" curriculum that is top-down imposed and the actual writing of the curriculum by the teachers themselves.

   Split between the "watchdogs" (administrators) and the "house masters" (dept. chairs & teachers) vis à vis job foci.

   District curric, coord'r as decision maker re school's acceptance of research relationship.

2. Summarize the information you got (or failed to get) on each of the target questions you had for this contact.

Question	Information
History of dev. of innov'n teachers	Conceptualized by Curric., Coord'r, English Chairman & Assoc. Chairman; written by teachers in summer; revised by following summer with field testing data
School's org'l structure	Principal & admin'rs responsible for discipline; dept chairs are educ'l leaders
Demographics emphasis	Racial conflicts in late 60's; 60% black stud. pop.; heavy on discipline & on keeping out non-district students slipping in from Chicago
Teachers' response to innov'n	Rigid, structured, etc. at first; now,they say they like it/ NEEDS EXPLORATION
Research access	Very good; only restriction: teachers not required to cooperate

3. Anything else that struck you as salient, interesting, illuminating or important in this contact?

   Thoroughness of the innov'n's development and training.

   Its embeddedness in the district's curriculum, as planned and executed by the district curriculum coordinator.

   The initial resistance to its high prescriptiveness (as reported by users) as contrasted with their current acceptance and approval of it (again, as reported by users).

4. What new (or remaining) target questions do you have in considering the next contact with this site?

   How do users really perceive the innov'n? If they do indeed embrace it, what accounts for the change from early resistance?

   Nature and amount of networking among users of innov'n.

   Information on "stubborn" math teachers whose ideas weren't heard initially—who are they? Situation particulars? Resolution?

   Follow-up on English teacher Reilly's "fall from the chairmanship."

   Follow a team through a day of rotation, planning, etc.

   CONCERN: The consequences of eating school cafeteria food two days per week for the next four or five months . . .

Stop

Source: Miles and Huberman (1994:10, Figure 4.1). Reprinted with permission from SAGE Publications, Inc.

TABLE 11.1 ■ Coding Example

Text	Code
He cares about me. He has never told me but he does.	Sense of self-worth
He's always been there for me, even when my parents were not. He's one of the few things that I hold as a constant in my life. So it's nice.	Stability
I really feel comfortable around him.	"comfortable"

Source: Saldaña, Johnny. (2021:6). *The Coding Manual for Qualitative Researchers*. Reprinted with permission from SAGE Publications, Inc.

analysis, tests them as the analysis proceeds, and then searches for new insights (Bingham and Witkowsky 2022). Anderson (2003) recounts how his conceptualization of social stratification at Jelly's Bar developed over a long period:

> I could see the social pyramid, how certain guys would group themselves and say in effect, "I'm here and you're there." . . . I made sense of these crowds [initially] as the "respectables," the "nonrespectables," and the "near-respectables." . . . Inside, such nonrespectables might sit on the crates, but if a respectable came along and wanted to sit there, the lower-status person would have to move. (pp. 225–226)

But this initial conceptualization changed with experience, as Anderson realized the participants themselves used other terms to differentiate social status: *winehead, hoodlum,* and *regular* (Anderson 2003:230). What did they mean by these terms? The regulars basically valued "decency." They associated decency with conventionality but also with "working for a living" or having a "visible means of support" (Anderson 2003:231). In this way, Anderson progressively refined his concept as he gained experience in the setting.

Howard Becker's (1958) study of medical students provides a classic illustration of this iterative process of conceptual organization:

> When we first heard medical students apply the term "crock" to patients, we made an effort to learn precisely what they meant by it. We found, through interviewing students about cases both they and the observer had seen, that the term referred in a derogatory way to patients with many subjective symptoms but no discernible physical pathology. Subsequent observations indicated that this usage was a regular feature of student behavior and thus that we should attempt to incorporate this fact into our model of student-patient behavior. The derogatory character of the term suggested in particular that we investigate the reasons students disliked these patients. We found that this dislike was related to what we discovered to be the students' perspective on medical school: the view that they were in school to get experience in recognizing and treating those common diseases most likely to be encountered in general practice. "Crocks," presumably having no disease, could furnish no such experience. We were thus led to specify connections between the student-patient relationship and the student's view of the purpose of this professional education. Questions concerning the genesis of this perspective led to discoveries about the organization of the student body and communication among students, phenomena which we had been assigning to another [segment of the larger theoretical model being developed]. Since "crocks" were also disliked because they gave the student no opportunity to assume medical responsibility, we were able to connect this aspect of the student-patient relationship with still another tentative model of the

value system and hierarchical organization of the school, in which medical responsibility plays an important role. (p. 658)

This excerpt shows how the researcher first was alerted to a concept by observations in the field, then refined his understanding of this concept by investigating its meaning. By observing the concept's frequency of use, he came to realize its importance. Then he incorporated the concept into an explanatory model of student–patient relationships.

A well-designed chart, or **matrix**, can facilitate the coding and categorization process. Table 11.2 shows an example of a coding form designed by Miles and Huberman (1994:93–95)

TABLE 11.2 ■ Example of a Checklist Matrix

Presence of Supporting Conditions		
Condition	**For Users**	**For Administrators**
Commitment	*Strong*—("wanted to make it work")	*Weak* at building level Prime movers in central office committed; others not
Understanding	*Basic* for teacher ("felt I could do it, but I just wasn't sure how") *Absent* for aide ("didn't understand how we were going to get all this")	*Absent* at building level and among staff *Basic* for 2 prime movers ("got all the help we needed from developer") *Absent* for other central office staff
Materials	*Inadequate:* ordered late, puzzling ("different from anything I ever used"), discarded	NA
Front-end training	"*Sketchy*" for teacher ("it all happened so quickly"); no demo class *None* for aide ("totally unprepared. I had to learn along with the children")	Prime movers in central office had training at developer site; none for others
Skills	*Weak-adequate* for teacher "*None*" for aide	One prime mover (Robeson) skilled in substance; others unskilled
Ongoing in-service	*None,* except for monthly committee meeting; no substitute funds	None
Planning, coordination time	*None*; both users on other tasks during day; lab tightly scheduled, no free time	None
Provisions for debugging	*None* systematized; spontaneous work done by users during summer	None
School admin. support	Adequate	NA
Central admin. support	*Very strong* on part of prime movers	Building admin. only acting on basis of central office commitment
Relevant prior experience	*Strong* and useful in both cases: had done individualized instruction, worked with low achievers, but aide had no diagnostic experience	*Present* and useful in central office, esp. Robeson (specialist)

Source: Miles and Huberman (1994:10, Table 5.2). Reprinted with permission from SAGE Publications, Inc.

to represent the extent to which teachers and teachers' aides ("users") and administrators at a school gave evidence of various supporting conditions that indicate preparedness for a new reading program. The matrix condenses data into simple categories, reflects further analysis of the data to identify degree of support, and provides a multidimensional summary that will facilitate subsequent, more intensive analysis. Direct quotes still impart some of the flavor of the original text.

Codes categorized in this way can then be condensed further into themes: a phrase or sentence that reflects a topic emerging from some of the codes and categories in the data (Saldana and Omasta 2022:305; Richards 2022). Gradually, the volume of text is reduced. As much as one half to two thirds of the total record may be summarized and deleted, with the rest subject to more intensive analysis. Ultimately, no more than 4% of excerpts are likely to be included in a published article. Saldana (2021:28) offers pithy advice: "Bottom line: Code smart, not hard. Code only the data that relate to your research questions of interest."

Examination and Display of Relationships

Examining relationships is the centerpiece of the analytic process because it allows the researcher to move from simple descriptions of the people and settings to explanations of why things happened as they did with those people in those settings. The process of examining relationships can be captured in a matrix that shows how different concepts are connected or perhaps what causes are linked with what effects.

Table 11.3 displays a matrix used to capture the relationship between the extent to which stakeholders in a new program had something important at stake in the program and the researcher's estimate of their favorability toward the program. Each cell of the matrix was to be filled in with a summary of an illustrative case study. In other matrix analyses, quotes might be included in the cells to represent the opinions of these different stakeholders, or the number of cases of each type might appear in the cells. The possibilities are almost endless. Keeping this approach in mind will generate many fruitful ideas for structuring a qualitative data analysis.

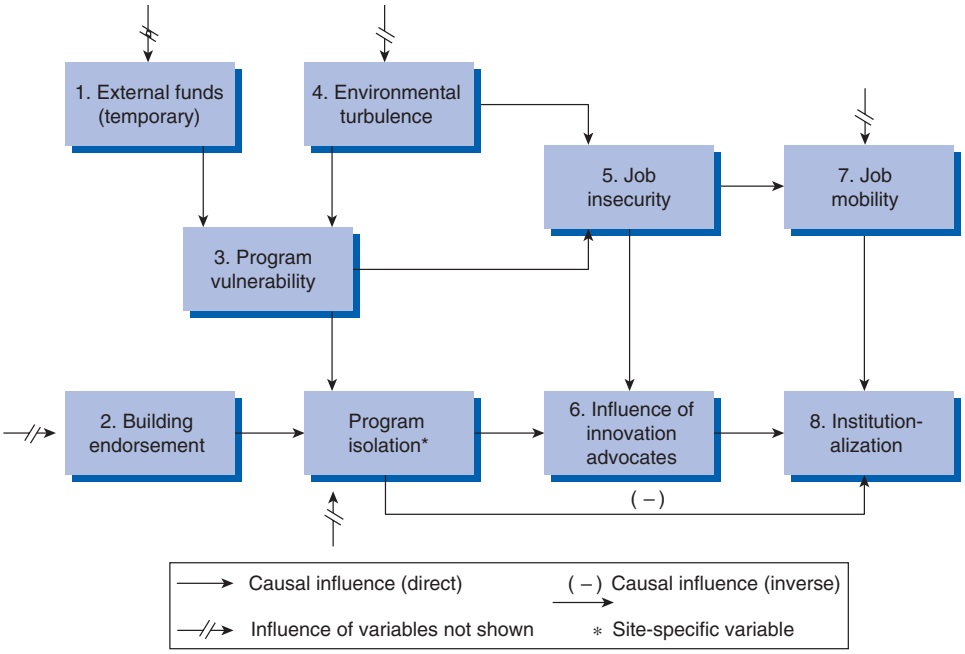
Simple relationships that are identified with a matrix, like that shown in Table 11.3, can be examined and then extended to create a more complex causal model. Such a model represents the multiple relationships between the constructs identified in a qualitative analysis as important for explaining some outcome. A great deal of analysis must precede the construction of such a model, with careful attention to identification of important variables and the evidence that suggests connections between them. Figure 11.4 provides an example of these connections from a study of the implementation of a school program.

TABLE 11.3 ■ Coding Form for Relationships: Stakeholders' Stakes

	Favorability Toward Program		
Importance	**Favorable**	**Neutral or Unknown**	**Antagonistic**
High			
Moderate			
Low			

Source: Patton (2002).

FIGURE 11.4 ■ Example of a Causal Network Model



Source: Miles and Huberman (1994:10, Figure 6.5). Reprinted with permission from SAGE Publications, Inc.

Corroboration and Legitimization of Conclusions

No set standards exist for evaluating the validity, or *authenticity*, of conclusions in a qualitative study, but the need to carefully consider the evidence and methods on which conclusions are based is just as great as with other types of research. Individual items of information can be assessed in terms of at least three criteria (Becker 1958):

1. *How credible was the informant?* Were statements made by someone with whom the researcher had a relationship of trust or by someone the researcher had just met? Did the informant have a reason to lie? If the statements do not seem to be trustworthy as indicators of actual events, can they at least be used to help understand the informant's perspective?
2. *Were statements made in response to the researcher's questions, or were they spontaneous?* Spontaneous statements are more likely to indicate what would have been said had the researcher not been present.
3. *How does the presence or absence of the researcher or the researcher's informant influence the actions and statements of other group members?* Reactivity to being observed can never be ruled out as a possible explanation for some directly observed social phenomenon. However, if the researcher carefully compares what the informant says goes on when the researcher is not present, what the researcher observes directly, and what other group members say about their normal practices, the extent of reactivity can be assessed to some extent. (pp. 654–656)

A qualitative researcher's conclusions should also be assessed by their ability to provide a credible explanation for some aspect of social life. That explanation should capture group members' **tacit knowledge** of the social processes that were observed, not just their verbal statements about these processes. Tacit knowledge—"the largely unarticulated, contextual understanding that is often manifested in nods, silences, humor, and naughty nuances"—is reflected in participants' actions as well as their words and in what they fail to state but nonetheless feel deeply and even take for granted (Altheide and Johnson 1994:492–493). These features are evident in William Foote Whyte's (1955) analysis of Cornerville social patterns:

> The corner-gang structure arises out of the habitual association of the members over a long period of time. The nuclei of most gangs can be traced back to early boyhood. . . . Home plays a very small role in the group activities of the corner boy. . . . The life of the corner boy proceeds along regular and narrowly circumscribed channels. . . . Out of [social interaction within the group] arises a system of mutual obligations which is fundamental to group cohesion. . . . The code of the corner boy requires him to help his friends when he can and to refrain from doing anything to harm them. When life in the group runs smoothly, the obligations binding members to one another are not explicitly recognized. (pp. 255–257)

Reflection on the Researcher's Role

Confidence in the conclusions from a field research study is also strengthened by an honest and informative account about how the researcher interacted with subjects in the field, what problems they encountered, and how these problems were or were not resolved. Such a "natural history" of the development of the evidence enables others to evaluate the findings and reflects the constructivist philosophy that guides many qualitative researchers (see Chapter 1). Such an account is important primarily because of the evolving and variable nature of field research: To an important extent, the researcher "makes up" the method in the context of a particular investigation rather than applying standard procedures that are specified before the investigation begins.

Barrie Thorne (1993) provides a good example of this final element of the analysis:

> Many of my observations concern the workings of gender categories in social life. For example, I trace the evocation of gender in the organization of everyday interactions, and the shift from boys and girls as loose aggregations to "the boys" and "the girls" as self-aware, gender-based groups. In writing about these processes, I discovered that different angles of vision lurk within seemingly simple choices of language. How, for example, should one describe a group of children? A phrase like "six girls and three boys were chasing by the tires" already assumes the relevance of gender. An alternative description of the same event—"nine fourth-graders were chasing by the tires"—emphasizes age and downplays gender. Although I found no tidy solutions, I have tried to be thoughtful about such choices. . . . After several months of observing at Oceanside, I realized that my field notes were peppered with the words "child" and "children," but that the children themselves rarely used the term. "What do they call themselves?" I badgered in an entry in my field notes. The answer it turned out, is that children use the same practices as adults. They refer to one another by using given names ("Sally," "Jack") or language specific to a given context ("that guy on first base"). They rarely have occasion to use age-generic terms. But when pressed to locate themselves in an age-based way, my informants used "kids" rather than "children." (pp. 8–9)

Qualitative data analysts, more often than quantitative researchers, display real sensitivity to how a social situation or process is interpreted from a particular background and set of values and not simply based on the situation itself (Altheide and Johnson 1994). Researchers are only human, after all, and must rely on their own senses and process all information through their own minds. By reporting how and why they think they did what they did, they can help others determine whether, or how, the researchers' perspectives influenced their conclusions. "There should be clear 'tracks' indicating the attempt [to show the hand of the ethnographer] has been made" (Altheide and Johnson 1994:493).

Anderson's (2003) memoir about the Jelly's Bar research illustrates the type of "tracks" that an ethnographer makes as well as how the researcher can describe those tracks. Anderson acknowledges that his tracks began as a child:

> While growing up in the segregated black community of South Bend, from an early age, I was curious about the goings-on in the neighborhood, particularly the streets and more particularly the corner taverns where my uncles and my dad would go to hang out and drink. . . . Hence, my selection of a field setting was a matter of my background, intuition, reason, and a little bit of luck. (pp. 217–218)

After starting to observe at Jelly's, Anderson's (2003) tracks led to Herman:

> After spending a couple of weeks at Jelly's, I met Herman. I felt that our meeting marked an important step. We would come to know each other well . . . something of an informal leader at Jelly's. . . . We were becoming friends. . . . He seemed to genuinely like me, and he was one person I could feel comfortable with. (pp. 218–219)

So we learn that Anderson's observations were to be shaped, in part, by Herman's perspective, but we also find out that Anderson maintained some engagement with fellow students. This contact outside the bar helped shape his analysis: "By relating my experiences to my fellow students, I began to develop a coherent perspective, or a 'story' of the place that complemented the accounts I had detailed in my accumulating field notes" (Anderson 2003:220).

In this way, Anderson explains that the outcome of his analysis of qualitative data resulted, in part, from the way in which he "played his role" as a researcher and participant, not just from the setting itself.

ALTERNATIVES IN QUALITATIVE DATA ANALYSIS

The qualitative data analyst can choose from many interesting alternative approaches. Of course, the research question under investigation should shape the selection of an analytic approach, but the researcher's preferences and experiences also will inevitably have an important influence on the method chosen. The alternative approaches I present here—grounded theory, abductive analysis, case-oriented understanding, conversation analysis, narrative analysis, and qualitative comparative analysis—give you a good sense of the different possibilities (Patton 2002).

Grounded Theory

Theory development occurs continually in qualitative data analysis (Coffey and Atkinson 1996:23). Many qualitative researchers use a method of developing theory during their analysis that is termed **grounded theory**, which involves building up inductively a systematic theory that is *grounded* in, or based on, the observations. The grounded theorist first summarizes observations into conceptual categories and then tests the coherence of these categories directly in the research

setting with more observations. Over time, as the researcher refines and links the conceptual categories, a theory evolves (Glaser and Strauss 1967; Huberman and Miles 1994:436).

As observation, interviewing, and reflection continue, grounded theory researchers refine their definitions of problems and concepts and select indicators. They can then check the frequency and distribution of phenomena: How many people made a particular type of comment? How often did social interaction lead to arguments? Social system models may then be developed that specify the relationships between different phenomena. These models are modified as researchers gain experience in the setting. For the final analysis, the researchers check their models carefully against their notes and make a concerted attempt to discover negative evidence that might suggest the model is incorrect.

Levitt, Swanger, and Butler (2008:435) used a systematic grounded method of analysis to understand the perspective of male perpetrators of violence on female victims. Research participants were recruited from programs the courts used in Memphis to assess and treat perpetrators who admitted to having physically abused a female intimate partner. All program participants were of low socioeconomic status, but in other respects, Levitt and her colleagues (2008:436) sought to recruit a diverse sample.

The researchers (Levitt et al. 2008:437–438) began the analysis of their interview transcripts by dividing them into "meaning units"—"segments of texts that each contain one main idea"—and labeling these units with terms like those used by participants. The researchers then compared these labels and combined them into larger descriptive categories. This process continued until they had combined all the meaning units into seven different clusters. Table 11.4 gives an

TABLE 11.4 ■ Clusters and Categories in a Grounded Theory Analysis

Clusters (endorsement)	Categories (endorsement)
1. The arrest incident is a hurdle or a test from god that I alone have to deal with, although the responsibility for the abuse was not all my own. (10)	**1.** If alcohol or drugs had not been in the picture, we wouldn't have come to blows: Substance use is thought to increase the rate of IPV (2) **2.** I don't want to get involved in conflict because I don't want to deal with its consequences (9) **3.** Joint responsibility in conflict depends on who did more fighting (8) **4.** How women cause IPV: Being treated as a child through nagging and being disrespected (5)
2. Passive avoidance and withdrawal from conflict is the best way to prevent aggression and to please god. (10)	**1.** DV thought to be "cured" by passively attending classes and learning anger management (6) **2.** Religious interventions have been vague or guilt producing, we need explicit advice and aren't getting it (9) **3.** Intimate partner violence can be stopped by cutting off relationships, but this can be a painful experience (5) **4.** Should resolve conflict to create harmony and avoid depression—but conflict may increase as a result (10)

Source: Levitt, H. M., Todd-Swanger, R., and Butler, J. B. "Male Perpetrators' Perspectives on Intimate Partner Violence, Religion, and Masculinity." *Sex Roles: A Journal of Research, 58*, 435–448. Copyright © 2007, Springer Science + Business Media, LLC. Reprinted with permission.

example of two of their clusters and the four categories of meaning units combined within each (Levitt et al. 2008:439).

With procedures such as these, the grounded theory approach develops general concepts from careful review of text or other qualitative materials and can then suggest plausible relationships between these concepts.

Abductive Analysis

Some qualitative researchers question grounded theory's assumption that it is possible to build theory from observations that "speak for themselves" and have developed an alternative approach called **abductive analysis**. Iddo Tavory and Stefan Timmermans (2014:15) define abductive analysis as a more balanced approach that seeks unexpected and puzzling observations, like grounded theory, but interprets these observations in relation to theories that we already know. The analytic process is then more of a back-and-forth "ongoing conversation in which both scholars and research subjects take part, challenging our theorizations and suggesting new ones" (Tavory and Timmermans 2014:125, 130).

The difference between abductive analysis and grounded theory is in part a difference in how we view the generation of theory in relation to data: Is it possible to approach data about the world that we experience with a "blank slate" and then figure out what is going on in the setting we have studied (grounded theory) or not (abductive analysis)? But the difference also has to do with how we think of the social role of the qualitative data analyst. In classical ethnography, the lone researcher goes "into the field" to gain experience without preconceptions and then records and analyzes their observations to make sense of that field. In abductive analysis, the researcher prepares for gathering observations by reading a broad range of theories and, after collecting observations, engages actively with the scholars and participants seeking to understand such settings (Tavory and Timmermans 2014:125, 131).

You don't need to think of abductive analysis as an entirely different approach to qualitative data analysis. Instead, recognize the problem of trying to approach qualitative data collection without preconceptions and consider orienting yourself to a setting for investigation by reading relevant theories of related social processes.

Case-Oriented Understanding

Like many qualitative approaches, a **case-oriented understanding** attempts to understand a phenomenon from the standpoint of the participants. The case-oriented understanding method reflects an interpretive research philosophy that is not geared to identifying causes but provides a different way to explain social phenomena. For example, Constance Fischer and Frederick Wertz (2002) constructed such an explanation of the effect of being criminally victimized. They first recounted crime victims' stories and then identified common themes in these stories:

> Their explanation began with a description of what they termed the process of "living routinely" *before the crime: "he/she . . . feels that the defended against crime could never happen to him/her.*" "I said, 'nah, you've got to be kidding.'" (pp. 288–289, emphasis in original)

> In a second stage, "being disrupted," the victim copes with the discovered crime and fears worse outcomes: "You imagine the worst when it's happening. . . . I just kept thinking my baby's upstairs." In a later stage, "reintegrating," the victim begins to assimilate the violation by taking some protective action: "But I clean out my purse now since then and I leave very little of that kind of stuff in there." (p. 289)

Finally, when the victim is "going on," he or she reflects on the changes the crime produced: "I don't think it made me stronger. It made me smarter." (p. 290)

You can see how Fischer and Wertz (2002:288–290) constructed an explanation of the effect of crime on its victims through this analysis of the process of responding to the experience. This effort to "understand" what happened in these cases gives us a much better sense of why things happened as they did.

RESEARCH IN THE NEWS

Who Are the Unvaccinated in America?

Dr. Laolu Fayanju, a family medicine doctor in Ohio, never expected that as many as one in five adult Americans would still be resisting vaccination for COVID-19 more than six months after the vaccines were first available. Dozens of interviews by *New York Times* reporters with unvaccinated Americans in 17 states identified a wide range of reasons for the decisions of those remaining unvaccinated. Fears, conspiracy theories, worries about safety, skepticism of the pharmaceutical industry and of the federal government each added to the numbers. An 85-year-old hospital worker questioned, "How do they know that it will kill the virus, and if it's really good for humans?" A landscaper claimed, "My immune system could fight it." A San Francisco firefighter reported, "I don't believe in mandates of any kind."

For Further Thought?

1. What predispositions and biases would you bring to the analysis of vaccine hesitancy? How could you plan your analysis of qualitative data from vaccine resisters to prevent your conclusions from being determined mostly by your own feelings rather than by the data you are analyzing? Can you identify a potential "critical friend"?
2. How would you approach analyzing interview data from people who have resisted vaccination? What steps would you take to ensure you captured the meanings they gave to their experiences? What evidence would you look for as you sought to construct a grounded theory about vaccine resistance topic?

News source: Bosman, Julie, Jan Hoffman, Margot Sanger-Katz, and Tim Arango. 2021. "Who Are the Unvaccinated in America." *The New York Times*, August 1.

Conversation Analysis

Conversation analysis is a specific qualitative method for analyzing the sequential organization and details of conversation. It focuses on how reality is constructed rather than on what it is. From this perspective, detailed analysis of conversational interaction is important because conversation is "sociological bedrock": a form of social organization through which the work of. . . institutions such as the economy, the polity, the family, socialization, etc." is accomplished (Schegloff 1996:4). It is how "we conduct the ordinary affairs of our lives" (Drew 2005:74).

Three premises guide conversation analysis (Gubrium and Holstein 2000:492):

1. Interaction is sequentially organized, and talk can be analyzed in terms of the process of social interaction rather than in terms of motives or social status.
2. Talk, as a process of social interaction, is contextually oriented—it is both shaped by interaction and creates the social context of that interaction.
3. These processes are involved in all social interaction, so no interactive details are irrelevant to understanding it.

Consider these premises as you read the following excerpt from Elizabeth Stokoe's (2006:479–480) analysis of the relevance of gender categories to "talk-in-interaction." The dialogue in Figure 11.5 is between four first-year British psychology students who must write up a description of some photographs of people. Stokoe incorporates stills from the video recording of the interaction into her analysis of both the talk and embodied conduct in interaction. In typical conversation analysis style, the text is broken up into brief segments that capture shifts in meaning, changes in the speaker, pauses, nonspeech utterances and nonverbal actions, and emphases.

Here, in part, is how Stokoe (2006) analyzes this conversation:

> In order to meet the task demands, one member of the group must write down their ideas. Barney's question at the start of the sequence, "is somebody scribing" is taken up after a reformulation: "who's writin' it." Note that, through a variety of strategies, members of the group manage their responses such that they do not have to take on the role of scribe. At line 05, Neil's "Oh yhe: ah." [the colon indicates prolongation of the phrase (Heritage n.d.) rather than a request for action, and his subsequent nomination of Kay directs the role away from himself. . . . At line 08, Neil nominates Kay, his pointing gesture working in aggregate with the talk to accomplish the action ("She wants to do it."), whilst also attributing agency to Kay for taking up the role. A gloss [interpretation] might be "Secretaries in general are female, you're female, so you in particular are our secretary." (p. 481)

Does Stokoe's analysis help you see how the social interaction reinforces the link of "woman" and "secretary"?

Narrative Analysis

Narrative methods use interviews and sometimes documents or observations to "follow participants down their trails" (Riessman 2008:24). Unlike conversation analysis, which focuses attention on moment-by-moment interchange, narrative analysis seeks to put together the "big picture" about experiences or events as the participants understand them. **Narrative analysis** focuses on "the story itself" and seeks to preserve the integrity of personal biographies or a series of events that cannot adequately be understood in terms of their discrete elements (Riessman 2008:23–27). The analysis will often begin with a reconstruction of the chronology of a life or series of events and then fill in the meanings attached to experiences throughout the chronology (Fawcett and Pockett 2015:59–62). Narrative thus "displays the goals and intentions of human actors; it makes individuals, cultures, societies, and historical epochs comprehensible as wholes" (Richardson 1995:200). The coding strategy for a narrative analysis revolves around reading the stories and classifying them into general patterns.

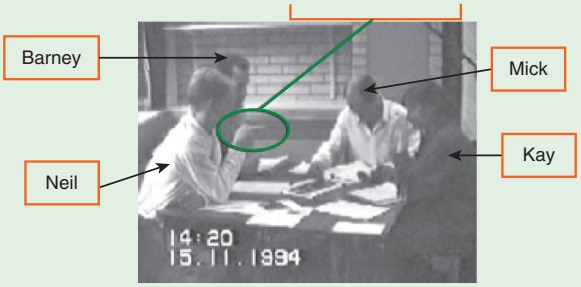
In an analysis of conflicts among high school students, Morrill and his colleagues (2000:534) read through 254 conflict narratives written by the ninth graders they studied and found four different types of stories:

1. *Action tales*, in which the author represents themself and others as acting within the parameters of taken-for-granted assumptions about what is expected for particular roles among peers.
2. *Expressive tales*, in which the author focuses on strong, negative emotional responses to someone who has wronged them.
3. *Moral tales*, in which the author recounts explicit norms that shaped their behavior in the story and influenced the behavior of others.
4. *Rational tales*, in which the author represents themself as a rational decision maker navigating through the events of the story.

FIGURE 11.5 ■ Conversation Analysis, Including Pictures

1. UT-23

01	N:		D' you reckon she's an instructor then.
02	N:		(0.2)
03	N:		Of some sort,
04	B:	→	Is somebody scribing. Who's writin' it. =
05	N:		=Oh yhe:ah.
06			(0.8)
07	M:		Well you can't [read my] =
08	N:		[((pointing to K)) She wants to do it.]



09	M:		=writin' once I've [wri:tten it.]
10	K:		[.hehhhh]
11	N:		We:ll secretary an' female.
12		→	(0.3)
13	K:		.Hh heh heh heh ((nodding, picks up pen and paper))

14			(0.4)
15	M:		It's uh::,
16	K:		Y

Source: Stokoe, Elizabeth. 2006. "On Ethnomethodology, Feminism, and the Analysis of Categorical Reference to Gender in Talk-In-Interaction." *The Sociological Review* 54:467–494.

Coding reliability was checked through a discussion between the two primary coders, who found their classifications agreed for a large percentage of the stories.

What Morrill et al. (2000) termed an *action tale* unfolds,

> in matter-of-fact tones kindled by dramatic tensions that begin with a disruption of the quotidian order of everyday routines. A shove, a bump, a look . . . triggers a response. . . . Authors of action tales typically organize their plots as linear streams of events as they move briskly through the story's scenes. . . . (p. 536)

Morrill et al. (2000) classify the following narrative as a *moral tale*, in which the students "explicitly tell about their moral reasoning, often referring to how normative commitments shape their decision-making" (p. 542):

> I . . . got into a fight because I wasn't allowed into the basketball game. I was being harassed by the captains that wouldn't pick me and also many of the players. The same type of things had happened almost every day where they called me bad words so I decided to teach the ring leader a lesson. I've never been in a fight before but I realized that sometimes you have to make a stand against the people that constantly hurt you, especially emotionally. I hit him in the face a couple of times and I got [the] respect I finally deserved. (pp. 545–546)

Morrill et al. (2000:553) summarize their classification of the youth narratives in a simple table that highlights the frequency of each type of narrative and the characteristics associated with each of them (see Table 11.5). How does such an analysis contribute to our understanding of youth violence? Morrill et al. (2000) first emphasize their narratives "suggest that consciousness of conflict among youths—like that among adults—is not a singular entity but comprises a rich and diverse range of perspectives" (p. 551).

TABLE 11.5 ■ Summary Comparison of Youth Narratives*

Representation of	Action Tales (*N* = 144)	Moral Tales (*N* = 51)	Expressive Tales (*N* = 35)	Rational Tales (*N* = 24)
Bases of everyday conflict	Disruption of everyday routines & expectations	Normative violation	Emotional provocation	Goal obstruction
Decision making	Intuitive	Principled stand	Sensual	Calculative choice
Conflict handling	Confrontational	Ritualistic	Cathartic	Deliberative
Physical violence†	In 44% (*N* = 67)	In 27% (*N* = 16)	In 49% (*N* = 20)	In 29% (*N* = 7)
Adults in youth conflict control	Invisible or background	Sources of rules	Agents of repression	Institutions of social control

*Total *N* = 254.

†Percentages based on the number of stories in each category.

Source: Morrill et al. 2000:553, Table 1, Copyright 2000. Reprinted with permission of Blackwell Publishing Ltd.

Theorizing inductively, Morrill et al. (2000:553–554) then attempt to explain why action tales were much more common than were the more adult-oriented normative, rational, or emotionally expressive tales. One possibility is Carol Gilligan's (1988) theory of moral development, which suggests that younger students are likely to limit themselves to the simpler action tales that "concentrate on taken-for-granted assumptions of their peer and wider cultures, rather than on more self-consciously reflective interpretation and evaluation" (Morrill et al. 2000:554). More generally, Morrill et al. (2000) argue, "We can begin to think of the building blocks of cultures as different narrative styles in which various aspects of reality are accentuated, constituted, or challenged, just as others are deemphasized or silenced" (p. 556).

In this way, Morrill et al.'s (2000) narrative analysis allowed an understanding of youth conflict to emerge from the youths' own stories while informing our understanding of broader social theories and processes.

Narrative inquiry can also be a much more active process, involving purposeful sampling of participants in a setting and then interviews that the researcher uses to construct a narrative about their lives that may even be shared and refined with feedback from the participants (Edmonds and Kennedy 2016:163). In this way, Magnus Kilger (2017) interviewed elite youth sports players in Sweden to learn how they constructed "success stories." After recording 53 interviews during training sessions and matches, he read and interpreted each one, creating a classification scheme that included four types of narratives: *the humble story*—"we trained a bit" and "it just moved on"; *the hard work story*—"I'm fighting quite a lot and I'm [sic] have a strong desire"; *the natural talent story*—"I've always been, like, sort of the best player in the team"; and *the superstar story*—"I have [a] great number of very strong qualities . . . also an extreme speed. . . . I'm very, very fast." In this way Kilger develops our understanding of how the players explain their success within their life stories. Narrative analysis can also use documents and observations and focus more attention on how stories are constructed rather than on the resulting narrative (Hyvärinen 2008:452).

Qualitative Comparative Analysis

Daniel Cress and David Snow (2000) collected qualitative data from about 15homeless social movement organizations (SMOs) in eight cities i. In each of these cities, Cress and Snow used a snowball sampling strategy to identify the homeless SMOs and the various supporters, antagonists, and significant organizational bystanders with whom they interacted. Cress and Snow then gathered information from representatives of these organizations, including churches, other activist organizations, police departments, mayors' offices, service providers, federal agencies, and, of course, the SMOs themselves.

To answer their research questions, Cress and Snow (2000) needed to operationalize each of the various conditions they believed might affect movement outcomes, using coding procedures that were much more systematic than those often employed in qualitative research. For example, Cress and Snow defined "sympathetic allies" operationally as

> the presence of one or more city council members who were supportive of local homeless mobilization. This was demonstrated by attending homeless SMO meetings and rallies and by taking initiatives to city agencies on behalf of the SMO. (Seven of the 14 SMOs had such allies.) (p. 1078)

Cress and Snow (2000) also chose a structured method of analysis, qualitative comparative analysis (QCA), to assess how the various conditions influenced SMO outcomes. This procedure

identifies the combination of factors that had to be present across multiple cases to produce a particular outcome (Ragin 1987). Cress and Snow (2000) explain why QCA was appropriate for their analysis:

> QCA . . . is conjunctural in its logic, examining the various ways in which specified factors interact and combine with one another to yield particular outcomes. This increases the prospect of discerning diversity and identifying different pathways that lead to an outcome of interest and thus makes this mode of analysis especially applicable to situations with complex patterns of interaction among the specified conditions. (p. 1079)

Table 11.6 summarizes the results of much of Cress and Snow's (2000) analysis. It shows that homeless SMOs that were coded as organizationally viable used disruptive tactics, had sympathetic political allies, and presented a coherent diagnosis and program in response to the problem they were protesting and were very likely to achieve all four valued outcomes: (1) representation, (2) resources, (3) protection of basic rights, and (4) some form of tangible relief. Some other combinations of the conditions were associated with increased likelihood of achieving some valued outcomes, but most of these alternatives less frequently had positive effects.

TABLE 11.6 ■ Multiple Pathways to Outcomes and Level of Impact

Pathways	Outcomes	Impact
1. VIABLE * DISRUPT * ALLIES * DIAG * PROG.....	Representation, Resources, Rights, and Relief	Very strong
2. VIABLE * disrupt * CITY * DIAG * PROG.....	Representation and Rights	Strong
3. VIABLE * ALLIES * CITY * DIAG * PROG.....	Resources and Relief	Moderate
4. viable * DISRUPT * allies * diag * PROG.....	Relief	Weak
5. viable * allies * city * diag * PRO G.....	Relief	Weak
6. viable * disrupt * ALLIES * CITY * diag * prog.....	Resources	Weak

Note: Uppercase letters indicate presence of condition and lowercase letters indicate the absence of a condition. Conditions not in the equation are considered irrelevant. Multiplication signs (*) are read as "and."

Source: Cress and Snow (2000:1097, Table 6). Copyright © 2000 The University of Chicago Press. Reprinted with permission.

The qualitative textual data on which the codes were based indicate how particular combinations of conditions exerted their influence. For example, one set of conditions that increased the likelihood of achieving increased protection of basic rights for homeless persons included avoiding disruptive tactics in cities that were more responsive to the SMOs.

Combining Qualitative Methods

Qualitative researchers often combine one or more of these methods within one analysis. Elif Kale-Lostuvali (2007) enriched his research on the effects of the İzmit earthquake in Turkey

by using a combination of qualitative methodologies—including participant observation and intensive interviewing.

One important concept that emerged from both the observations and the interviews was the distinction between a *magˇdur* (sufferer) and a *depremzade* (son of the earthquake). This was a critical distinction because a *magˇdur* was seen as deserving of government assistance, but a *depremzade* was considered to be taking advantage of the situation for personal gain. Kale-Lostuvali (2007) drew on both interviews and participant observation to develop an understanding of this complex concept:

> A prominent narrative that was told and retold in various versions all the time in the disaster area elaborated the contrast between *magˇdur* (sufferer; that is, the truly needy) and *depremzades* (sons of the earthquake) on the other. The *magˇdur* (sufferers) were the deserving recipients of the aid that was being distributed. However, they (1) were in great pain and could not pursue what they needed; or (2) were proud and could not speak of their need; or (3) were humble, always grateful for the little they got, and were certainly not after material gains; or (4) were characterized by a combination of the preceding. And because of these characteristics, they had not been receiving their rightful share of the aid and resources. In contrast, *depremzades* (sons of the earthquake) were people who took advantage of the situation. (p. 755)

Analyses based on combining different qualitative methods in this way can yield a richer understanding of the social context under investigation.

VISUAL SOCIOLOGY

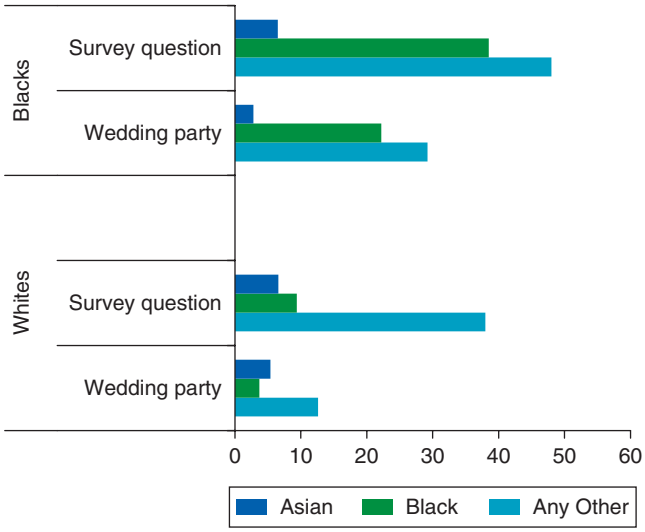
In the 1970s, qualitative researchers Howard Becker (1974) and Erving Goffman (1979) drew attention to the value of visual images in understanding social patterns, but in more recent decades, the availability of photos and videos has exploded because of the ease of taking them with smartphones and posting them on the Internet (Rainie et al. 2012). By 2021, 95 million photos and videos were being shared on Instagram every day, more than 1 billion videos were being viewed on TikTok, and more than 1 billion hours of videos were being watched on YouTube (Heitman 2021). As a result, increasing numbers of social scientists are collecting and analyzing visual representations of social life, and **visual sociology** has become a growth industry.

You have already seen in this chapter how Stokoe's conversation analysis of "gender talk" (2006) was enriched by her analysis of photographs. Continuous video recordings can help researchers unravel sequences of events and identify nonverbal expressions of feelings (Heath and Luff 2008:501). You also see later in this chapter how Peter St. Jean (2007)) used systematic coding of videotaped observations to measure the extent of disorder in St. Louis neighborhoods.

In an innovative, visually based approach to studying interracial friendship patterns, Brent Berry (2006) sampled wedding photos that had been posted on the web. Berry compared the rate of different-race members of wedding parties to the prevalence of different-race friends reported in representative surveys. As you can see in Figure 11.6, answers to survey questions create the impression that interracial friendships are considerably more common than is indicated by the wedding party photos.

An analysis by Eric Margolis (2004) of photographic representations of American Indian boarding schools provides a good example of how the analysis of visual materials created by others can help us understand cultural change. On the left side of Figure 11.7 is a picture taken in 1886 of members of the Chiricahua Apaches who had just arrived at the Carlisle Indian

FIGURE 11.6 ■ Interracial Friendships in Wedding Party Photos and in Responses to Survey Questions, by Race



Source: Based on Berry, Brent. 2006. "Friends for Better or for Worse: Interracial Friendship in the United States as Seen Through Wedding Party Photos." *Demography* 43:491–510.

FIGURE 11.7 ■ Pictures of Chiricahua Apache Children Before and After Starting Carlisle Indian School, Carlisle, Pennsylvania, 1886

Source: Margolis, Eric. 2004. "Looking at Discipline, Looking at Labour: Photographic Representations of Indian Boarding Schools." *Visual Studies* 19:72–96.

School in Carlisle, Pennsylvania. The school was run by Captain Richard Pratt, who, like many Americans in that period, felt tribal societies were communistic, indolent, dirty, and ignorant, whereas Western civilization was industrious and individualistic. So Pratt set out to acculturate American Indians to the dominant culture. The second picture shows the result: the same group of Apaches looking like Europeans, not Native Americans—dressed in standard uniforms, with standard haircuts, and with more standard posture.

Many other pictures display the same type of transformation. Are these pictures each "worth a thousand words"? They capture the ideology of the school management, but we can be less certain they document accurately the "before and after" status of the students. Pratt "consciously

used photography to represent the boarding school mission as successful" (Margolis 2004:79). Although he clearly tried to ensure a high degree of conformity, there were accusations the contrasting images were exaggerated to overemphasize the change (Margolis 2004:78). Reality was being constructed, not just depicted, in these photographs. Visual sociologists must always consider the purposes for which pictures were created and the extent to which people consciously posed for the occasion (Tinkler 2013:31). Even more important, visual representations must be analyzed in the context of the associated text and other indications of the social context of the photos (Pink 2021:40). "It is important that visual researchers make every effort to situate the subject of their research and their specific take on it in its broader context, both visually and verbally" (Pauwels 2010:564).

The ease with which video can be captured with smartphones makes "walking methods" more feasible (Pink 2021:98) as well as easier to integrate with other methods. U.K. researchers Nick Emmel and Andrew Clark (2011) discuss how photographs collected in "walkarounds" enriched their understanding of the social setting they studied (see Figure 11.8):

> The research is situated in one geographical location or fieldsite. Periodically we walked through this field along a set pathway taking photographs. . . . The research is conducted in a geographical place covering around 1.5 mile^2 (circa 2.5 km^2) with a mixed population. Relatively affluent students live in close proximity to one of the most deprived populations in England. . . . Within this socially heterogeneous geographical context our research explores, among other aims, the ways different social groups create, maintain, dissemble, and experience, social networks over time and across space. We each use the photographs we take on the walk as an adjunct to the other methods we are using in the research. . . . . They contribute to and facilitate an interpretation of place, which in turn provides a more complete account of the place and space in which we are doing research . . . how this analytical process happens. The panorama could be analysed at face value as

FIGURE 11.8 ■ A Playground in the Fieldsite of Emmel and Clark

Source: Emmel, Nick, and Andrew Clark. 2011. "Learning to Use Visual Methodologies in Our Research: A Dialogue Between Two Researchers." *Forum: Qualitative Social Research* 12(1): Article 36, http://nbn-resolving.de/urn: nbn:de:0114-fqs1101360.

> an empty play area; perhaps supporting ideas about the out-migration of families (a common theme discussed by some resident groups). . . .Subsequent questioning about play spaces in the area however, reveals a range of alternative explanations for under-use. For example, conversational interviews with young people reveal a more nuanced geography of play and socialisation in the area; informal discussion with a local official suggest [sic] infrastructural problems with this particular space, while analysis of the recent history of this play space hints at a more political explanation for its existence and apparent under-use. This means that I do not analyse the images alone (that is, as a discrete data set); but rather alongside other methods. . . .Finally, . . . I use the walkaround method as a way of formulating new questions to ask of participants in the other methods. In some respects, it is the making of the photograph (deciding whether, and what, to photograph and why), rather than the image itself, that is more analytically revealing (pp. 2–3, 8–10).

The ease with which video can be captured with smartphones is also increasing the use of video in ethnographic projects. **Video ethnography** combines the ethnographic process of engaging in ongoing social processes in a defined social setting with the visual sociologist's attention to the nonverbal aspects of the social world. Although professional video ethnographic projects require use of a dedicated video camera, such as a handycam with a "shotgun" microphone mounted on it that records sound directly in front of the camera (Shrum and Scott 2017:52–53), you can develop some skill in the method with just your smartphone.

With video gear in hand, a video ethnographer needs to consider the boundaries around the social processes of interest and what to film to capture those processes. It's best seen as a component of an ethnographic study rather than as a method in itself, for the video ethnographer has to "be there" when key interactions happen (Shrum and Scott 2017:28). Of course, recording with a camera injects a new element into most social processes, but the widespread use of cameras in smartphones means that the presence of an ethnographer may be more disruptive than the fact of using a camera. A useful final product—a video that "tells a story"—will only emerge after extensive editing. The video editing process will require specialized practice (Shrum and Scott 2017:93).

Photo voice is a method of using photography to engage research participants in explaining how they have made sense of their social worlds. The researcher directing a photo voice project distributes cameras to research participants and invites them to take pictures of their surroundings or everyday activities. The participants then meet with the researcher to present their pictures and discuss their meaning. In this way, researchers learn more about the participants' social world as they see it and react to it.

Jyoti Saksena and Shannon L. McMorrow (2020) combined Photovoice with intensive interviews in a study of 16 adult women refugees from the Democratic Republic of Congo who had been living in Indianapolis for no more than two years. In group meetings, the researchers asked the women to take photos in the community and their homes of things that made them happy, sad, or surprised about life in the United States. The photos were then discussed in meetings, and the women were asked to check the captions researchers used to summarize the photos. Figure 11.9 includes one of these pictures and the associated caption.

SYSTEMATIC OBSERVATION

Observations can be made in a more systematic, quantitative design that allows systematic comparisons and more confident generalizations. A researcher using **systematic observation** develops a standard form on which to record variation within the observed setting in terms of variables of

FIGURE 11.9 ■ Picture and Caption in Photo Voice Project

Source: Saksena, Jyotika, and Shannon L. McMorrow. 2020. "Through Their Eyes: A Photovoice and Interview Exploration of Integration Experiences of Congolese Refugee Women in Indianapolis." *Journal of International Migration and Integration* 21(2):529–49.

interest. Such variables might include the frequency of some behavior(s), the particular people observed, the weather or other environmental conditions, and the number and state of repair of physical structures. In some systematic observation studies, records will be obtained from a random sample of places or times.

Peter St. Jean (2007) used systematic social observation as well as resident surveys, participant observation, and in-depth interviews with residents and offenders to study crime in a St. Louis neighborhood. Using video cameras, St. Jean recorded neighborhood physical and social appearances from a van driven along neighborhood streets. Pictures were then coded for the presence of neighborhood disorder (see Figure 11.10).

The systematic observations give us much greater confidence in the measurement of relative neighborhood disorder than we would have from unstructured descriptive reports or from responses of residents to survey questions. Interviews with residents and participant observation helped identify the reasons offenders chose particular locations when deciding where to commit crimes.

PARTICIPATORY ACTION RESEARCH

Whyte (1991) urged social researchers to engage with research participants throughout the research process, which he argued would lead to more valid and insightful conclusions. He formalized this recommendation into an approach he termed **participatory action research (PAR)**. As the name implies, this approach encourages social researchers to get "out of the academic rut" and bring values into the research process (p. 285). Since Whyte's early call for this type

FIGURE 11.10 ■ One Building in St. Jean's (2007) Study

Source: © Peter K. B. St. Jean. Reprinted with permission.

of research, PAR has become increasingly popular in disciplines ranging from public health to social work, as well as sociology (McIntyre 2008; Minkler 2000). Participatory action research is not itself a qualitative method, but PAR projects tend to use qualitative methods, which are more accessible to members of the lay public and which typically involve some of the same activities as in PAR: engaging with individuals in their natural settings and listening to them in their own words.

In PAR, also termed *community-based participatory research* (CBPR), the researcher involves as active participants some members of the setting studied. "The goal of CBPR is to create an effective translational process that will increase bidirectional connections between academics and the communities that they study" (Hacker 2013:2). For example, many academic studies have found employee participation is associated with job satisfaction but not with employee productivity. After some discussion about this finding with employees and managers, Whyte realized researchers had been using a general concept of employee participation that did not distinguish those aspects of participation that were most likely to influence productivity (pp. 278–279). For example, occasional employee participation in company meetings had not been distinguished from ongoing employee participation in and control of production decisions. When these and other concepts were defined more precisely, it became clear that employee participation in production decisions had substantially increased overall productivity, whereas simple meeting attendance had not. This discovery would not have occurred without the active involvement of company employees in planning the research.

Those who engage in PAR projects are making a commitment "to listen to, learn from, solicit and respect the contributions of, and share power, information, and credit for

accomplishments with the groups that they are trying [to] learn about and help" (Horowitz, Robinson, and Seifer 2009:2634). The emphasis on developing a partnership with the community is reflected in the characteristics of the method presented in a leading health journal (see Figure 11.11). Each characteristic (except "emphasis on multiple determinants of health") identifies a feature of the researcher's relationship with community members.

FIGURE 11.11 ■ Characteristics of Community-Based Participatory Research (CBPR)

Community members and researchers contribute equally and in all phases of research.

Trust, collaboration, shared decision making, and shared ownership of the research; findings and knowledge benefit all partners.

Researchers and community members recognize each other's expertise in bidirectional, colearning process.

Balance rigorous research and tangible community action.

Embrace skills, strength, resources, and assets of local individuals and organizations.

Community recognized as a unit of identity.

Emphasis on multiple determinants of health.

Partners commit to long-term research relationships.

Core elements include local capacity building, system development, empowerment, and sustainability.

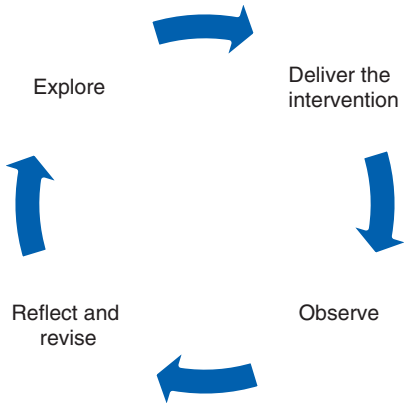
Source: Horowitz, Carol R., Mimsie Robinson, and Sarena Seifer. 2009. "Community-Based Participatory Research From the Margin to the Mainstream: Are Researchers Prepared?" *Circulation* May 19, 2009. Table 1, Characteristics of CBPR:2634. Reprinted with permission of Wolters Kluwer Health, LWW.

PAR can bring researchers into closer contact with participants in the research setting through groups that discuss and plan research steps and then take steps to implement research findings. For this reason, PAR is "particularly useful for emergent problems for which community partners are in search of solutions but evidence is lacking" (Hacker 2013:8). Stephen Kemmis and Robin McTaggart (2005:563–568) summarize the key steps in the process of conducting a PAR project as creating "a spiral of self-reflecting cycles":

- Planning a change
- Acting and observing the process and consequences of the change
- Reflecting on these processes and consequences
- Replanning
- Acting and observing again

In contrast to the formal reporting of results at the end of a research project, these cycles make research reporting an ongoing part of the research process. "Community partners can deepen the interpretation process once results are available, as they are intimately familiar with the context and meaning" (Hacker 2013:8). Community partners may also work with academic researchers to make changes in the community reflecting the research findings. Publication of results is only part of this cyclical research approach (see Figure 11.12).

FIGURE 11.12 ■ The Cyclical Action Research Approach



Source: Edmonds, W. Alex, and Thomas D. Kennedy. 2016. An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods, 2nd ed, Thousand Oaks, CA: SAGE.

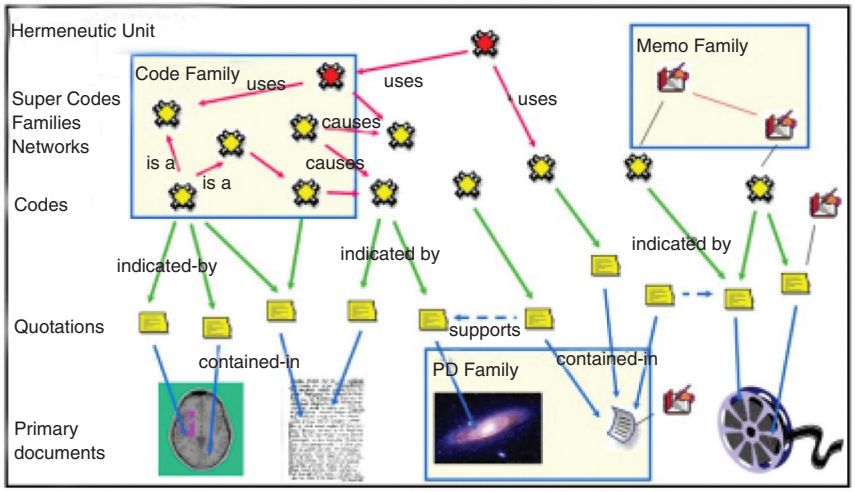
Karen Hacker, while at Harvard University and the Institute for Community Health in Cambridge, Massachusetts, collaborated with community partners in response to a public health emergency in the adjacent town of Somerville (Hacker 2013:8–10; Hacker et al. 2008). After a series of youth suicides and overdoses from 2000 to 2005, a PAR community coalition was formed with members from mental health service providers, school leaders, police, and community parents. After reviewing multiple statistics, the coalition concluded the deaths represented a considerable increase over previous years. However, when mental health professionals attempted to interview family members of adolescents who had died by suicide to investigate the background to the suicides, they were rebuffed; in contrast, family members were willing to talk at length with PAR members from their community. The PAR team was then able to map the relationships between the adolescents. The process of using the results of this research to respond to the suicides included a candlelight vigil, a speak-out against substance abuse, the provision of crisis counseling, and programs to support families and educate the community. Subsequently, the suicide rate dropped back to its pre-2000 level.

COMPUTER-ASSISTED QUALITATIVE DATA ANALYSIS

The analysis process can be enhanced in various ways by using a computer. Programs designed for qualitative data can speed up the analysis process, make it easier for researchers to experiment with different codes, test different hypotheses about relationships, and facilitate diagrams of emerging theories and preparation of research reports (Coffey and Atkinson 1996; Richards and Richards 1994). The steps involved in **computer-assisted qualitative data analysis** parallel those used traditionally to analyze text such as notes, documents, or interview transcripts: preparation, coding, analysis, and reporting. We use two of the most popular programs to illustrate these steps: QSR NVivo and ATLAS.ti.

Text preparation begins with typing or scanning text in a word processor or with NVivo directly into the program's rich text editor. NVivo will create or import a rich text file. These programs now allow multiple types of files, including pictures and videos as well as text. Figure 11.13 displays the different file types and how they are connected in the organization of a project (a "hermeneutic unit") with ATLAS.ti.

FIGURE 11.13 ■ File Types and Unit Structure in ATLAS.ti

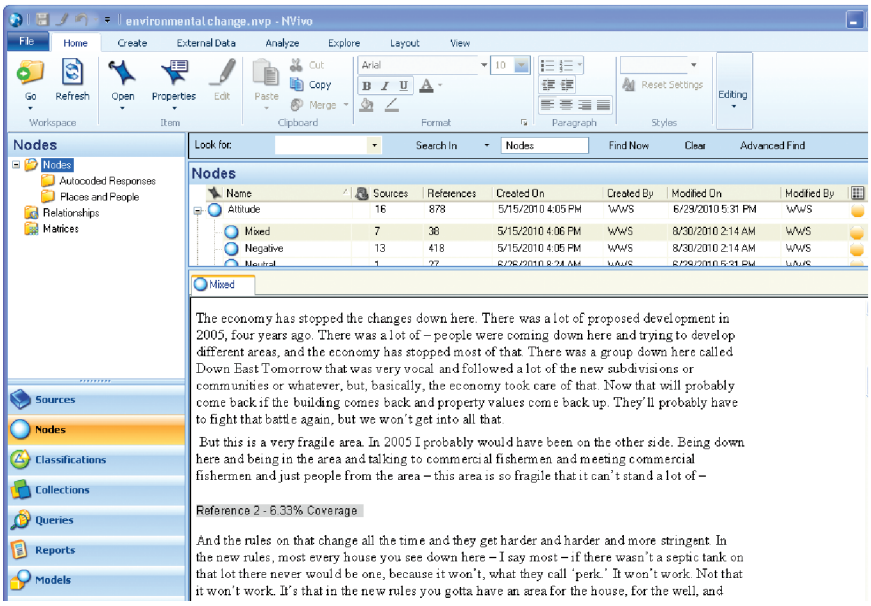


Source: Muhr and Friese (2004:29).

Coding the text involves categorizing text segments. This is the foundation of much qualitative analysis. Each program allows you to assign a code to any segment of text (in NVivo, you drag through the characters to select them). You can make up codes as you go through a document and assign codes you have already developed to text segments. Figure 11.14 shows the screen that appears in NVivo at the coding stage. The program also allows you to examine the coded text "in context"—embedded in its place in the original document.

In qualitative data analysis, coding is not a one-time-only or one-code-only procedure. Each program allows you to be inductive and holistic in your coding: You can revise codes as you go along,

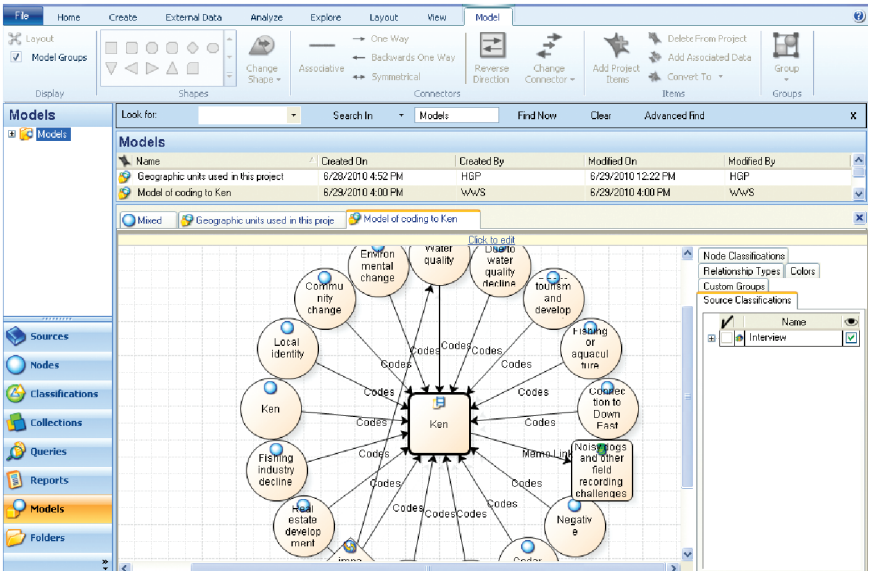
FIGURE 11.14 ■ NVivo Coding Stage


environmental change.nvp - NVivo
Nodes
Nodes
Sources
Nodes
Classifications
Collections
Queries
Reports
Models
Mixed
The economy has stopped the changes down here. There was a lot of proposed development in 2005, four years ago. There was a lot of – people were coming down here and trying to develop different areas, and the economy has stopped most of that. There was a group down here called Down East Tomorrow that was very vocal and followed a lot of the new subdivisions or communities or whatever, but, basically, the economy took care of that. Now that will probably come back if the building comes back and property values come back up. They'll probably have to fight that battle again, but we won't get into all that.
But this is a very fragile area. In 2005 I probably would have been on the other side. Being down here and being in the area and talking to commercial fishermen and meeting commercial fishermen and just people from the area – this area is so fragile that it can't stand a lot of –
Reference 2 - 6.33% Coverage
And the rules on that change all the time and they get harder and harder and more stringent. In the new rules, most every house you see down here – I say most – if there wasn't a septic tank on that lot there never would be one, because it won't, what they call 'perk.' It won't work. Not that it won't work. It's that in the new rules you gotta have an area for the house, for the well, and

assign multiple codes to text segments, and link your own comments ("memos") to text segments. You can work "live" with the coded text to alter coding or create new, more subtle categories. You can also place hyperlinks to other documents in the project or to any multimedia files outside it.

Analysis focuses on reviewing cases or text segments with similar codes and examining relationships between different codes. You may decide to combine codes into larger concepts. You may specify additional codes to capture more fully the variation between cases. You can test hypotheses about relationships between codes and develop more free-form models (see Figure 11.15). You can specify combinations of codes that identify cases you want to examine.

FIGURE 11.15 ■ A Free-Form Model in NVivo



Reports from each program can include text to illustrate the cases, codes, and relationships that you specify. You can also generate counts of code frequencies and then import these counts into a statistical program for quantitative analysis. However, the many types of analyses and reports that can be developed with qualitative analysis software do not lessen the need for a careful evaluation of the quality of the data on which conclusions are based.

The first step in preparing to analyze qualitative data collected online is to consider what the data represent. Are the data textual, verbal, visual, or some mix of those types? Do the communications represent contacts that are one-on-one, one-to-many, or many-to-many? Table 11.7 gives a few examples of each. Your plans also have to take into account whether the material you will study existed without any researcher action (extant), were elicited in response to questions posted by you or another researcher (elicited), or were generated through an interactive process that might have involved reactions to simulations or vignettes (enacted) (Salmons 2016:32). Now that you have defined the type of data you have to work with, you can apply the techniques of qualitative data analysis you learned earlier in this chapter.

ETHICS IN QUALITATIVE DATA ANALYSIS

The qualitative data analyst is never far from ethical issues and dilemmas. Data collection should not begin unless the researcher has a plan that others see as likely to produce useful knowledge. Relations developed with research participants and other stakeholders to facilitate data

TABLE 11.7 ■ Internet Communication Options

Mode	One-to-One	One-to-Many	Many-to-Many
Text	E-mail, text message	Blogs, e-mail list	Social networking sites, discussion forums
Verbal	Voice over Internet protocol call (VoIP)	Podcast	VoIP conference calls
Visual	Image attachments or links, avatars	Interaction as avatar representation	Virtual worlds
Mixed	Video chat	Webinar	Massively open online games

Source: Salmons, Janet E. 2016. *Doing Qualitative Research Online*. Thousand Oaks, CA: SAGE.

collection should also be used to keep these groups informed about research progress and findings. Research participants should be encouraged to speak out about emerging study findings (Lincoln 2009:154–155). Decisions to reproduce photos and other visual materials must be considered in light of privacy and copyright issues. Throughout the analytic process, the analyst must consider how the findings will be used and how participants in the setting will react. The need to minimize harm requires attention, even after data collection has concluded.

As I noted in Chapter 10, some Indigenous peoples have established rules for outside researchers to minimize harm to their community and preserve their autonomy. These rules may require collaboration with an Indigenous researcher, collective approval of admission to their culture, and a review of all research products before publication. The primary ethical commitment is to the welfare of the community as a whole and the preservation of their culture rather than to the rights of individuals (Lincoln 2009:162–163).

Participatory action research projects create special challenges because the researcher is likely to feel divided loyalties to the community and the university. Even though it may be difficult for researchers to disagree with the interpretations of community members about the implications of research findings, they may feel a different interpretation is warranted (Sieber and Tolich 2013:94, 102–104). Qualitative researchers need to be sensitive to the potential for these problems and respond flexibly and respectfully to the concerns of research participants (Hacker 2013:109–110). Allowing community leaders and members to comment on the researcher's interpretations is one way to lessen feelings of exclusion. However, it is also important to recognize "the community" does not refer to a homogeneous singular entity. As Martin Levinson (2010) explains, "When dealing with human communities, groups are actually made up of individuals, and the interests of the group do not necessarily converge with those of the individuals that constitute it" (p. 205).

Qualitative researchers should also take care to consider harm that might occur along all points in the research process. Staff who transcribe interviews or focus group transcripts may not seem a part of the formal research process—and may even be independent contractors who are only contacted through the Internet—but they may react emotionally to some of the material in particularly charged interviews. In projects where transcripts contain emotionally charged material, those used as transcriptionists should be given advance warning, and providing connections to counselors should be considered (Sieber and Tolich 2013:175).

Miles and Huberman (1994:293–295) suggest specific questions that are particularly important during the process of data analysis. The next several sections describe these questions.

Privacy, Confidentiality, and Anonymity

"In what ways will the study intrude, come closer to people than they want? How will information be guarded? How identifiable are the individuals and organizations studied?" We have considered these issues already in the context of qualitative data collection, but they also must be a concern during the process of analysis. It can be difficult to present a rich description in a case study while not identifying the setting. It can be easy for participants in the study to identify each other in a qualitative description, even if outsiders cannot. The concerns are heightened when using visual data involving pictures of research participants or others (Harley and Langdon 2018).

I confronted the challenge of maintaining confidentiality in my analysis of ethnographic notes collected in the study of housing for homeless persons diagnosed with severe mental illness that I used in Chapter 10 to illustrate field notes. I explained the dilemma and the approach I took in the methodological appendix to *Homelessness, Housing, and Mental Illness* (Schutt 2011b):

> It is not possible to understand what happened in the [six] houses, or to consider possible explanations for particular events, without reading detailed descriptions of what people did and having at least some examples of what people said. But providing such details about specific identified group homes would often reveal who it was who was speaking or acting, at least to former housing staff and participants.
>
> In order to avoid providing intimate and possibly identifiable stories about participants, I have adhered throughout this book to three more specific guidelines. First, I have used arbitrary pseudonyms to refer to everyone whose name appears in the ethnographic notes that I include. Sometimes, I simply refer to individuals by their role, such as "tenant" or "house staff." Second, I have not maintained consistent pseudonyms for individual participants across chapters or even across different incidents that I describe within chapters. This greatly reduces the risk that even a former participant could determine another's identity by gradually being able to imagine the person as I describe different things they said or did. Third, I do not link the activities or statements I report to the individual houses. Since there were only seven different project group homes (six at any one time), it would be much easier for a former participant to link some actions or statements to another participant if the house in which they occurred were identified. However, every group home in our project experienced many of the same types of incidents and interchanges, so by not distinguishing the houses, I left the statements difficult to connect to particular individuals.
>
> Of course, this procedure for maintaining participant confidentiality has a cost for the analysis: I cannot "tell the story" of the development of particular houses with the ethnographic details about the houses. However, review of the notes from all the group homes indicates that the similarities in the issues they confronted far outweighed their differences. Moreover, our three ethnographers varied in their styles of observing, interviewing, and even note-taking, so that explicit comparisons of particular houses could be only partial. So maintaining the confidentiality of the individual group homes seems prudent from a methodological standpoint as well as essential from the standpoint of protecting the confidentiality of our research participants. (pp. 317–318)

In ethnographic and other participant studies where initial access is negotiated with community leaders or with groups of participants, qualitative researchers should discuss with participants the approach that will be taken to protect privacy and maintain confidentiality. Selected participants should also be asked to review reports or other products before their public release to gauge the extent to which they feel privacy has been appropriately preserved. Research with photographs that identify individuals raises special ethical concerns. Although legal standards are evolving, it is important not to violate an individual's expectations of privacy in any setting and to seek informed consent for the use of images when privacy is expected (Tinkler 2013:196–198).

Analysis of data collected online creates additional challenges. Guidelines for obtaining voluntary consent were presented in Chapter 10 for data elicited from online participants. When extant data are analyzed, the terms of membership or registration, if any, should be reviewed for the website(s) studied. If it seems that members or participants have been told that their remarks are confidential or that access to the site is restricted, further efforts are needed to secure participants' consent for use of their statements. It is also a good idea to check with the site moderator, if any, and to review the participants' interactions and statements for evidence of expectations of privacy. If in doubt, posting a statement about the research and requesting comment on the intended use of the data is good practice (Salmons 2016:86–91).

Intervention and Advocacy

"What do I do when I see harmful, illegal, or wrongful behavior on the part of others during a study? Should I speak for anyone's interests besides my own? If so, whose interests do I advocate?" Maintaining what is called *guilty knowledge* may force the researcher to suppress some parts of the analysis so as not to disclose the wrongful behavior, but presenting "what really happened" in a report may prevent ongoing access and violate understandings with participants. The need for intervention and advocacy is more likely to be anticipated in PAR/CBPR projects because they involve ongoing engagement with community partners who are likely to have an action orientation (Hacker 2013:101–104).

Research Integrity andQquality

"Is my study being conducted carefully, thoughtfully, and correctly in terms of some reasonable set of standards?" Real analyses have real consequences, so you owe it to yourself and those you study to adhere strictly to the analysis methods you believe will produce authentic, valid conclusions. Visual images that demean individuals or groups should not be included in publications (Tinkler 2013:197).

Ownership of Data and Conclusions

"Who owns my field notes and analyses: I, my organization, my funders? And once my reports are written, who controls their diffusion?" Of course, these concerns arise in any social research project, but the intimate involvement of the qualitative researcher with participants in the setting studied makes conflicts of interest between different stakeholders much more difficult to resolve. Working through the issues as they arise is essential. Mitch Duneier (1999:319–330) decided to end *Sidewalk*, his ethnography of New York City sidewalk book vendors, with an afterword by one of his key informants. Such approaches that allow participants access to conclusions in advance and the privilege to comment on them should be considered in relation to qualitative projects. The public availability of visual images on websites does not eliminate concerns about ownership. Copyright law in the United States as well as in the United Kingdom

and Australia provides copyright to content on the Internet as soon as it is uploaded, but there are disagreements about the requirement of informed consent before reproducing images from publicly accessible sites (Tinkler 2013:204–205). Researchers leading PAR/CBPR projects must work out data ownership agreements in advance of data collection to ensure there are no misunderstandings about retention of data and maintenance of confidentiality after the project ends (Hacker 2013:99–100).

Use and Misuse of Results

"Do I have an obligation to help my findings be used appropriately? What if they are used harmfully or wrongly?" It is prudent to develop understandings early in the project with all major stakeholders that specify what actions will be taken to encourage appropriate use of project results and to respond to what is considered misuse of these results. Visual researchers must also consider how participants will feel about their images appearing in publications in the future Wiles et al. (2012):

> People take part in our research, and they don't think in terms of publications arising years and years later. . . . So I think there are lots of problems, even when you have formally and legally the consent they have signed, because it refers to much earlier . . . she might have changed, it's a few years, she might feel very differently, it might remind her now of something very unpleasant. (Youth researcher, focus group 3:48)

Video ethnographers should always make sure participants know how their images and statements will be used and then avoid depictions that could cause harm to individuals. Consent should be recorded in signed "release and liability waiver statements," although filmed statements of permission for the filming may be appropriate in unplanned situations (Shrum and Scott 2017:121–122).

PAR/CBPR projects are designed to help solve local problems, but harm might also occur if results are not what were expected or if findings cast some elements of the community in an unfavorable light. These possibilities should be addressed as the analysis progresses and resolved before they are publicized (Hacker 2013:114–117).

CONCLUSIONS

The variety of approaches to qualitative data analysis makes it difficult to provide a consistent set of criteria for interpreting their quality. Norman Denzin's (2002:362–363) "interpretive criteria" are a good place to start. Denzin suggests that at the conclusion of their analyses, qualitative data analysts ask the following questions about the materials they have produced. Reviewing several of them will serve as a fitting summary for your understanding of the qualitative analysis process:

- Do they illuminate the phenomenon as lived experience? In other words, do the materials bring the setting alive in terms of the people in that setting?
- Are they based on thickly contextualized materials? We should expect thick descriptions that encompass the social setting studied.
- Are they historically and relationally grounded? There must be a sense of the passage of time between events and the presence of relationships between social actors.

- Are they processual and interactional? The researcher must have described the research process and their interactions within the setting.
- Do they engulf what is known about the phenomenon? This includes situating the analysis in the context of prior research and acknowledging the researcher's own orientation on first starting the investigation.

When an analysis of qualitative data is judged as successful in terms of these criteria, we can conclude that the goal of authenticity has been achieved.

As a research methodologist, you should be ready to use qualitative techniques, evaluate research findings in terms of these criteria, and mix and match specific analysis methods as required by the research problem to be investigated and the setting in which it is to be studied.

KEY TERMS

Abductive analysis (p. 404)
Case-oriented understanding (p. 404)
Computer-assisted qualitative data analysis (p. 418)
Conversation analysis (p. 405)
Emic focus (p. 389)
Etic focus (p. 389)
Grounded theory (p. 402)
Matrix (p. 398)
Narrative analysis (p. 406)
Participatory action research (PAR) (p. 415)
Photo voice (p. 414)
Progressive focusing (p. 389)
Qualitative comparative analysis (QCA) (p. 409)
Systematic observation (p. 414)
Tacit knowledge (p. 401)
Video ethnography (p. 414)
Visual sociology (p. 411)

HIGHLIGHTS

LO 11.1 Explain the meaning of an emic focus and of an etic focus in qualitative research.

- An emic focus occurs when a researcher represents a setting with the participants' terms and from their viewpoint. Most qualitative researchers have an emic focus.
- An etic focus occurs when a researcher represents a setting with their own terms and from their viewpoint. Most quantitative researchers have an etic focus.
- Qualitative data analysts seek to describe their textual data in ways that capture the setting or people who produced this text on their own terms rather than in terms of predefined measures and hypotheses.

LO 11.2 List the five different techniques that are shared by most approaches to qualitative data analysis

- Documentation of the data and the process of data collection
- Organization, categorization, and condensation of the data into concepts
- Examination and display of relationships between concepts
- Corroboration and legitimization of conclusions, by evaluating alternative explanations, disconfirming evidence, and searching for negative cases
- Reflection on the researcher's role

LO 11.3 Defend the value of using more than one method of analyzing qualitative data in a project.

- Alternative approaches to qualitative data analysis include grounded theory, abductive analysis, case-oriented understanding, conversation analysis, narrative analysis, ethnomethodology, and qualitative comparative analysis.
- Combining different qualitative methods can allow for better understandings of complex topics and yield a richer understanding of the social context under investigation.

LO 11.4 Identify changes in the social world that have led to the growth of visual sociology.

- In recent decades, the availability of photos and videos to be analyzed by visual sociologists has exploded because of the ease of taking them with smartphones and posting them on the Internet.
- The ease with which video can be captured with smartphones makes "walking methods," where a researcher walks through a fieldsite collecting images, more feasible.
- Today, video ethnographic projects can be completed on smartphones and no longer require the use of a dedicated video camera and microphone.
- The widespread use of cameras in smartphones means that the presence of an ethnographer may be more disruptive than the fact of using a camera.

LO 11.5 Argue for the strengths of a systematic observation technique.

- Systematic observation creates standardization of records; the development of a standard form allows the researcher to consistently record variation within the observed setting in terms of variables of interest.
- Systematic observation inspires much more confidence in measurements of variables of interest than unstructured descriptive reports or survey responses.
- Systematic observation allows for more systematic comparison between cases and greater generalizability.

LO 11.6 Appraise the strengths of participatory action research.

- Whyte (1991) argued that PAR would lead to more valid and insightful conclusions and get researchers "out of the academic rut."
- PAR can bring researchers into closer contact with participants in the research setting through groups that discuss and plan research steps and then take steps to implement research findings.
- Implementation of PAR, in particular incorporating input from members of the population being studied, can help solve emergent problems in a research study that do not have obvious solutions.
- Individuals may be more willing to speak to PAR members from their community than other types of individuals attempting to complete research.

LO 11.7 Recommend the ways in which computer-assisted qualitative data analysis can facilitate research.

- Computer-assisted qualitative data analysis allows a researcher to create, apply, and refine categories; trace linkages between concepts; and make comparisons between cases and events.

- Computer-assisted qualitative data analysis programs can speed up the analysis process, make it easier for researchers to experiment with different codes, test different hypotheses about relationships, and facilitate diagrams of emerging theories and the preparation of research reports.
- Computer-assisted qualitative data analysis programs allow you to be inductive and holistic in your coding: You can revise codes as you go along, assign multiple codes to text segments, and link your own comments ("memos") to text segments. You can work "live" with the coded text to alter coding or create new, more subtle categories. You can also place hyperlinks to other documents in the project or to any multimedia files outside it.
- Computer-assisted qualitative data analysis programs reports can include text to illustrate the cases, codes, and relationships that you specify. You can also generate counts of code frequencies and then import these counts into a statistical program for quantitative analysis.

LO 11.8 Review the following ethical issues that should be given special attention in qualitative data analysis: (1) privacy, confidentiality, and anonymity; (2) research integrity and quality; (3) ownership of data and conclusions; and (4) use and misuse of results.

- *Privacy, confidentiality, and anonymity:* It can be difficult to present a rich description in a case study while not identifying the setting. It can be easy for participants in the study to identify each other in a qualitative description, even if outsiders cannot. The concerns are heightened when using visual data involving pictures of research participants or others.
- *Research integrity and quality:* Real analyses have real consequences, so you owe it to yourself and those you study to adhere strictly to the analysis methods you believe will produce authentic, valid conclusions.
- *Ownership of data and conclusions:* the intimate involvement of the qualitative researcher with participants in the setting studied makes conflicts of interest between different stakeholders much more difficult to resolve.
- *Use and misuse of results*: It is prudent to develop understandings early in the project with all major stakeholders that specify what actions will be taken to encourage appropriate use of project results and to respond to what is considered misuse of these results. Visual researchers must also consider how participants will feel about their images appearing in publications in the future.43

DISCUSSION QUESTIONS

1. List the primary components of qualitative data analysis strategies. Compare and contrast each of these components with those relevant to quantitative data analysis. What are the similarities and differences? What differences do these make?
2. Does qualitative data analysis result in trustworthy results—in findings that achieve the goal of authenticity? Why would anyone question its use? What would you reply to the doubters?

3. Narrative analysis provides the "large picture" of how a life or event has unfolded, whereas conversation analysis focuses on the details of verbal interchange. When is each method most appropriate? How could one method add to another?

4. Ethnography (described in Chapter 10), grounded theory, and case-oriented understanding each refer to aspects of data analysis that are an inherent part of the qualitative approach. What do these approaches have in common? How do they differ? Can you identify elements of these three approaches in this chapter's examples of ethnomethodology, qualitative comparative analysis, visual sociology, and narrative analysis?

COMPUTER-ASSISTED QUALITATIVE DATA ANALYSIS

1. QSR NVivo © provides many tutorials for those learning to analyze data in NVivo. The first three give you an overview of the process: https://help-nv.qsrinternational.com/20/win/Content/tutorials/nvivo-tutorials.htm. After you review these three tutorials, identify some of the differences between the process of qualitative and quantitative data analysis.

2. Now you can pick and choose from 43 tutorials on specific NVivo capabilities that are available on YouTube. Review the list and watch two tutorials that interest you. How does qualitative analysis software facilitate the analysis process? Might it hinder the qualitative data analysis process in some ways? On balance, would you prefer to analyze qualitative data with a program like NVivo or on paper, using your notes and transcriptions? Explain your answers.

SECTION IV

COMPLEX SOCIAL RESEARCH DESIGNS

Chapter 12 Mixed Methods 430

Chapter 13 Evaluation and Policy Research 453

Chapter 14 Research Using Secondary Data and "Big" Data 488

Chapter 15 Research Using Historical and Comparative Data and Content Analysis 521

Chapter 16 Summarizing and Reporting Research 559

©iStockphoto.com/AtnoYdur

12 MIXED METHODS

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

12.1 Describe the development of mixed methods

12.2 Diagram mixed methods that differ in sequence and priority

12.3 Devise research studies that exemplify each major type of mixed method

12.4 Critique the strengths and limitations of each major type of mixed method

12.5 Identify ethical issues in at least one example of a project using each type of mixed method

The initial COVID-19 lockdown forced members of most families to spend more time together, confined at home, and so led to concerns among social workers and others about increasing rates of child abuse and partner violence. Initial reports seemed to confirm a rapid increase in domestic abuse worldwide as early as April 2020 (Taub 2020). Studies of previous natural and manmade disasters suggested such an increase would be expected (Seddighi et al. 2019). But by June 2020, reports of child abuse in New York City had dropped by half compared to the same period in 2019 (Stewart 2020). What was going on? Had reports of interpersonal violence dropped because child protection agencies were not able to visit people in their homes? Were abuse partners forced to stay at home preventing their victims from reporting abuse? Was violence within families increasing, decreasing, or remaining stable?

Bas Tierolf at the Verwey-Jonker Institute, with Edith Geurts at the Augeo Foundation and Majone Steketee at Erasmus University decided to find out what was going on during the lockdown among families in the Netherlands with histories of domestic abuse (Tierolf, Geurts, Steketee 2021). An ongoing longitudinal questionnaire-based study of vulnerable families provided before-and-after quantitative data about domestic abuse. But the complexities of the pandemic situation convinced them they shouldn't only rely on quantitative measures, so they added a qualitative study to their research design. In this way, they created a mixed-methods project.

Although Tierolf et al.'s (2021) research focused on just one aspect of the social impact of the pandemic in just one country, their study highlights a general problem: A single method may not represent adequately the social world's complexity that we are trying to understand. The goal of this chapter is to highlight the importance of this lesson and the need for mixed-methods research to overcome it. I will discuss the historical and philosophical foundation for mixed methods, illustrate the different ways that methods can be combined, and highlight some of the advantages and challenges of using multiple methods in one study. Although in other chapters I mention some investigations that combine different methods, what is distinctive about this chapter is its focus on research designs that have major qualitative and quantitative components as well as its attention to the challenges of combining methods.

HISTORY OF MIXED METHODS

Sociologists and other social scientists have long used multiple methods in their research, but only in recent decades have some focused attention on how best to combine qualitative and quantitative methods to better achieve research goals. As an example of the classic tradition, anthropologist Warner et al.'s (1960) classic *Social Class in America* used both qualitative and quantitative research methods to understand the social complexities of U.S. communities. Based largely on his pioneering field research in the 1930s and 1940s in the New England community of Newburyport—called "Yankee City" in his subsequent books—Warner showed how qualitative data about social position and activities could be converted into a quantitative measure that he termed the "Index of Status Characteristics" (I.S.C.). Here is an example of how Warner et al. (1960) shifted from qualitative description to quantitative assessment:

> The home of Mr. and Mrs. Henry Adams Breckenridge . . . has three stories and is topped by a captain's walk. . . . Large trees and a tall thick hedge . . . garden stretches one hundred yards . . . many old rose bushes. . . . The life and surroundings of Mrs. Henry Adams Breckenridge, old-family and upper-upper, . . . Her ratings of the characteristics of her I.S.C. give her a final score of 12, or perfect. . . . (pp. 243, 247)

Another classic example of the early use of **mixed methods** is represented in *Union Democracy*, sociologist Seymour Martin Lipset's study of the International Typographical Union (ITU) with Martin Trow and James Coleman (1956). After some preliminary investigations involving reading union literature and interviewing ITU members, Lipset began systematic qualitative research.

> Aside from long exploratory interviews with key informants in the union, members of the research team that was organized around the study familiarized themselves in every way possible with the actual political life of the union, attending union meetings, party caucuses, and chapel meetings, while paying particular attention to the events preceding the local union election held in May 1951. (Lipset et al. 1956:xiii).

The qualitative research "led to a sharpening and respecifying of hypotheses" and the decision to add quantitative methods.

> At this point it seemed that certain crucial aspects of the internal political process of the ITU could best be studied through survey research methods, and moreover, that such a study could be practically carried out among the members of the New York local. (Lipset et al. 1956:xiii–xiv)

Throughout *Union Democracy*, Lipset et al. (1956:470–492) analyzed both qualitative and quantitative data to identify and illustrate factors that influence political practices in the union. However, they did not discuss the ways in which their use of both quantitative and qualitative methods helped them to answer research questions or created particular challenges.

Donald Campbell and Donald Fiske (1959) were an early exception to this history of avoiding explicit discussion of mixed methods. They proposed that a "**multitrait–multimethod matrix**" in validity studies would allow greater confidence in the validity of measures when the results of measuring the same phenomenon with different methods converge and when the results of measuring different phenomena with the same method diverge. This article quickly became a classic, but there was little additional explicit attention to the value of mixed methods for the next two decades.

Triangulation was a more general term used at that time to refer to the use of multiple methods to study one research question (see pp. 124–125). The term suggests that a researcher can get a clearer picture of the social reality being studied by viewing it from several different perspectives. Each will have some liabilities in a specific research application, and all can benefit from a combination of one or more other methods (Brewer and Hunter 1989; Sechrest and Sidani 1995).

In 1981, Carolyn Needleman (1981) added a qualitative component to her quantitative investigation of a juvenile justice system in New York. Needleman used statistics to determine whether juvenile offenders were more likely to be sent to court—meaning they were being treated more harshly, she supposed—if they were older, if they were African American, if they were male, if they already had a record of infractions, and so on. But Needleman also sought to understand the processing of juvenile cases, and so she observed the probation officers in the intake office and interviewed them repeatedly over a two-year period.

What Needleman learned from her qualitative participant observation is surprising: Probation workers did not regard a referral to court as a harsher sanction than being diverted from court (Needleman 1981:248). In fact, the intake workers believed the court was too lenient in its decisions and instead felt their own "adjustment" effort during probation, with counseling, negotiation, restitution, and social service agencies was harsher and more effective (Needleman 1981:249):

> Considered toothless, useless and counterproductive, the court represents to the intake workers not a harsher sanction, but a dumping ground for the "failures" from the intake's alternative [probation]. (Needleman 1981:253)

Perhaps most troubling for research using case records, Needleman found that probation officers decided how to handle cases first and then created an official record that appeared to justify their decisions. Table 12.1 summarizes the differences Needleman's qualitative interviews identified between researchers' assumptions about juvenile court workers and the juvenile court workers' own assumptions.

TABLE 12.1 ■ Researchers' and Juvenile Court Workers' Discrepant Assumptions

Researcher Assumptions	Intake Worker Assumptions
● Being sent to court is a harsher sanction than diversion from court.	● Being sent to court often results in more lenient and less effective treatment.
● Screening involves judgments about individual juveniles.	● Screening centers on the juvenile's social situation.
● Official records accurately capture case facts.	● Records are manipulated to achieve the desired outcome.

Source: Needleman (1981:248–256).

Not until 1989 did John Brewer and Albert Hunter's book *Multi-Method Research* herald more systematic attention to this approach. Increasing interest in mixed methods reflected to some extent the increasingly complex research problems that social researchers were confronting and the demands of policy makers and others for sophisticated, convincing evidence. Since that time, mixed methods have been used more in applied fields than in academic disciplines: 16% of articles in nursing and education used mixed methods in a sample drawn in 2005, compared

with 6% of articles in sociology and psychology (Alise and Teddlie 2010:115). There is also evidence in many publications of growing acceptance of the value of qualitative methods and of recognition by many qualitative researchers of the need for more generalizable quantitative evidence (Alise and Teddlie 2010:115; Creswell and Plano Clark 2011:20–22).

The turn toward mixed methods in the late 1980s also reflected a decline in disagreements between qualitative and quantitative methodologists about the philosophies behind their methodological preferences. After this earlier period of conflict known as the "**paradigm wars**," social researchers were more willing to consider qualitative and quantitative methods as complementary (Alise and Teddlie 2010).

Marcus Weaver-Hightower (2013) explained the basis for his rejection of the paradigm wars approach in his report on a mixed-methods study of policy making about boys' education in Australia:

> Both qualitative and quantitative methods do have limitations for studying policy influence; qualitative methods can have difficulty establishing the extent of influence while quantitative methods can have difficulty providing the whys, hows, and so whats. Rather than succumb to paralysis from competing claims for methodological incompleteness, I used mixed methods to ameliorate each approach's limitations. (p. 6)

Ryan Brown and his colleagues (2013) at the RAND Corporation were also guided by this logic when they sought to learn why most health intervention strategies with homeless adults have had limited effect:

> Properly addressing such complexity and interdependency requires agile research strategies that can not only assess causal factors at multiple levels but also flexibly incorporate new information as it arises during the research process. This means enabling creative and productive conversation between qualitative and quantitative measurement and analytic modalities—a mixed-methods approach. (p. 329)

DIAGRAMMING MIXED METHODS

A researcher who plans to use both qualitative and quantitative research methods to best address a research problem is adopting a mixed-methods approach (Plano Clark and Ivankova 2016:57). However, the specific ways to "mix" the methods vary between research problems and settings. Qualitative methods may be used before or after quantitative methods; that is, the sequencing of the two approaches can differ. In addition, one method may be given priority in a project, or qualitative and quantitative methods may be given equal priority. Distinguishing the sequencing and priority of the qualitative and quantitative methods used in any mixed-methods project results in several different types. Although some mixed methods researchers use different labels for the types, and others believe different methods should just be combined as needed in a project without worrying about "types," I think you'll find that reading about the types will give you a better sense of the options (Plano Clark and Ivankova 2016:126–128).

Before discussing these types, it will help to learn some basic conventions for naming them:

- The primary method used in a mixed-methods project is written in all caps (QUAN or QUAL).
- The secondary method is written in lowercase letters (quan or qual).

- If both methods are given equal priority, they are both written in all caps.
- If one method is used before the other, the sequence is indicated with an arrow (QUAL→quan or Qual→QUAN or QUAN→QUAL, etc.).
- If two methods are used concurrently but one has priority, the secondary method is said to be "embedded" in the primary method. This is indicated as follows: QUAL(quan) or QUAN(qual).
- If two methods are used concurrently but they have equal priority, the relation between the two methods is indicated with a + symbol: QUAL+QUAN.

The different types of mixed methods that result from these distinctions in priority and sequence are represented in Table 12.2. Of course, these different types represent only research projects in which each method is used only once; more complex mixed-methods projects may involve combinations of these types, either at the same time or in sequence.

TABLE 12.2 ■ Types of Mixed Methods

		Priority	
		Prioritized	**Equal**
Sequence	**Sequential**	Staged Method Qual→QUAN Quan→QUAL QUAL→quan QUAN→qual	**Complex Research Program** **QUAL→QUAN** **QUAN→QUAL** **or other more complex combinations**
	Concurrent	**Embedded Method** **QUAL(quan)** **QUAN(qual)**	**Integrated Method** **QUAL+QUAN**

Some examples using the research projects I have already introduced will help to clarify the reasons for using these different types (Creswell and Plano Clark 2011:8; Morgan 2014:20). Needleman (1981) collected quantitative data from case records on dispositions in juvenile cases, but she realized she might misinterpret the meaning of these numbers unless she also understood the meaning behind the probation officers' recorded dispositions. She therefore embedded a qualitative study of probation officers within her larger quantitative study of case dispositions. This project can be represented as using a QUAN(qual) embedded method.

In their study of the International Typographical Union, Lipset and colleagues (1956) needed to identify the differences in orientations of the union's various social groups and to understand how these differences were expressed in union politics. For this purpose, they used an integrated method (QUAL+QUAN) in which both quantitative methods (a structured survey) and qualitative methods (in-depth interviews) played major roles. The Netherlands study of domestic violence during the pandemic by Tierolf et al. (2021) also used an integrated QUAL+QUAN design. In both their description of methods and their presentation of findings, Tierolf et al. (2021) give equal attention to the quantitative survey of families before and after the lockdown and the qualitative interviews with families and professionals. They then integrate their analysis of these findings in their discussion section.

You learned in the section on refining and testing questions in Chapter 8 that qualitative methods like focus groups and cognitive interviews can be used to refine questions for a structured quantitative survey: qua-->QUAN. In the section on organizing concepts inChapter 10, you saw that observations may be coded to create systematic quantitative data: QUAL-->quan. In these types of staged methods, one method has priority. In addition, the use of one method precedes the other one in time because the data collected with one method shape the data collected with the other method.

TYPES OF MIXED METHODS

In this section, I will discuss in more depth examples of research projects using integrated, staged, and embedded methods. When qualitative and quantitative methods are combined in these ways, researchers must make key decisions about how best to mix methods to achieve their study purposes. As you see how researchers make these decisions, you will also understand better the strengths and weaknesses of the specific methods.

Integrated Mixed-Methods Designs

In an **integrated mixed-methods design**, qualitative and quantitative methods are used concurrently and both are given equal importance. Findings produced from these methods are then integrated and compared during the analysis of project data. This is the QUAL+QUAN design. Susan A. McCarter (2009) extended prior research on juvenile justice processing with an integrated mixed-methods investigation of case processing and participant orientations in Virginia. Her hope was to use the results of the two methods to triangulate the study's findings; that is, to show that different methods lead to similar conclusions and so increase the credibility of her findings.

For her quantitative analysis, McCarter (2009) used data collected on 2,233 African American and Caucasian males in Virginia's juvenile justice system:

> The quantitative data set (n = 2,920) is a disproportionate, stratified, random sample of juvenile cases from all 35 Virginia Court Service Units (CSU) where each CSU was treated as a separate stratum. These data were collected by the Joint Legislative Audit and Review Commission (JLARC) in an examination of court processing and outcomes of delinquents and status offenders in Virginia.
>
> JLARC collected data on the juveniles' previous felonies; previous misdemeanors; previous violations of probation/parole; previous status offenses; recent criminal charges, intake action on those charges, pre-disposition(s) of those charges, court disposition(s) of those charges; and demographics such as sex, race, data of birth, CSU, and geotype (urban, suburban, rural). For a subset of these cases, data included information from the youth's social history, which required judicial request. (p. 535*)

In-depth interviews with 24 juvenile judges, attorneys, police officers, juveniles, and their families provided the qualitative data for the study (McCarter 2009):

> The juvenile justice personnel were from six Court Service Units across the state, including two urban, two suburban, two rural, two from Region I, two from Region II, and two from Region III. . . . Participants from each CSU were chosen to provide maximum diversity in perspectives and experiences, and thus varied by race, sex, and age; and the

> justice personnel also varied in length of employment, educational discipline and educational attainment. (p. 536) . . . The youth and their families were all selected from one Court Service Unit (CSU) located in an urban geotype with a population of approximately 250,000 (p. 536). The sample of youth and their family members was comprised of all male juveniles, five mothers and one father. Four of the six families were African American and two were Caucasian. (p. 540*)
>
> **McCarter, Susan A. 2009. "Legal and Extralegal Factors Affecting Minority Overrepresentation in Virginia's Juvenile Justice System: A Mixed-Method Study." Child and Adolescent Social Work Journal 26:533–544. Copyright © 2009, Springer. Reprinted with permission.*

The in-depth interviews included both open- and closed-ended questions. The open-ended responses were coded into categories that distinguished how participants perceived the role of race in the juvenile justice system (McCarter 2009:536). A direct connection with the quantitative findings was made in the interviews themselves:

> Respondents were read the quantitative findings from this study and then asked whether or not their experiences and/or perceptions of the juvenile justice system were congruent with the findings. They were also asked how commonly they believed instances of racial or ethnic bias occurred in Virginia. (McCarter 2009:540)

Comments made in response to this qualitative question supported the quantitative finding that race mattered in the juvenile justice system:

> Juvenile justice professionals as well as youth and their families cited racial bias by individual decision-makers and by the overall system, and noted that this bias was most likely to occur by the police during the Alleged Act or Informal Handling stages. However, although race was considered a factor, when compared to other factors, professionals did not think race played a dominant role in affecting a youth's treatment within the juvenile justice system. . . . Eighteen of the juvenile justice professionals stated that they felt a disparity [between processing of African American and white juveniles] existed, four did not feel that a disparity existed, and two indicated that they did not know. (McCarter 2009:540)

In this way, the qualitative and quantitative findings were integrated, and the study's key conclusion about race-based treatment was strengthened because it was based on triangulated identification (McCarter 2009:542)—supporting earlier conclusions of quantitative research by Dannefer and Schutt (1982).

In their integrated mixed-methods study of families who had been reported for possible domestic violence in the Netherlands, Tierolf et al. (2021) compared responses from 159 parents (and some children) who had completed questionnaires before the pandemic began to those of 87 parents who completed questionnaires soon after the pandemic lockdown began. They found no difference in the frequency of reports of violence before and after the lockdown. Among teenagers, feelings of emotional insecurity decreased during the lockdown. The qualitative study helped explain these findings. They conducted qualitative in-depth interviews by phone with 39 parents and 9 children from these families, who agreed to being interviewed, as well as with 13 professionals who worked in child protection agencies. The lockdown had increased stress in most families, but parents and children coped by going to their bedroom or going outside for a while, and everyone tried to be more considerate of each other due to the need for each

other during the pandemic. Some families reported things were calmer with fewer social workers in the house and teenagers experienced less school pressure. Professionals found new ways of staying in touch with the families. In general, the research drew attention to the importance of sources of resilience and to the different ways that different families experienced the lockdown.

Embedded Mixed-Methods Designs

Maria Testa, Jennifer Livingston, and Carol VanZile-Tamsen (2011) at the University at Buffalo's Research Institute on Addictions supplemented a quantitative study of violence against women on college campuses with a qualitative component. They explained that because violence against women is "a complex, multifaceted phenomenon, occurring within a social context that is influenced by gender norms, interpersonal relationships, and sexual scripts," "understanding of these experiences of violence is dependent on the subjective meaning for the woman and cannot easily be reduced to a checklist" (p. 237). This was an **embedded mixed-methods design**.

Victims' responses to structured questions indicated an association between alcohol and rape, but when victims elaborated on their experiences in qualitative interviews, their comments led to a new way of understanding this quantitative association. Although this association has often been interpreted as suggesting "impaired judgment" about consent by intoxicated victims, the women interviewed by Testa et al. (2011) all revealed they had had so much to drink they were unconscious or at least unable to speak at the time of the rape. Testa and her colleagues (2011) concluded the prevalence of this type of "incapacitated rape" required a new approach to the problem of violence against women:

> Qualitative analysis of our data has resulted in numerous "a-ha" types of insights that would not have been possible had we relied solely on quantitative data analysis (e.g., identification of incapacitated rape and sexual precedence, heterogeneity in the way that sexual assaults arise) and also helped us to understand puzzling quantitative observations. . . . These insights, in turn, led to testable, quantitative hypotheses that supported our qualitative findings, lending rigor and convergence to the process. We never could have anticipated what these insights would be and that is what is both scary and exhilarating about qualitative data analysis, particularly for a scientist who has relied on quantitative data analysis and a priori hypothesis testing. The lengthy process of reading, coding, rereading, interpreting, discussing, and synthesizing among two or more coders is undeniably a major investment of time. (p. 242)

Testa and her colleagues concluded that insights yielded by the qualitative data analysis fully justified the time-consuming process of reading and rereading interviews, coding text, and discussing and reinterpreting the codes (p. 245).

Staged Mixed-Methods Designs

Migration in search of employment creates family strains and lifestyle changes in countries around the world. This is nowhere as true as in contemporary China, where millions of young people have migrated from rural villages to industrial cities. Juan Zhong and Jeffrey Arnett (2014) sought to understand the impact of such migration on Chinese women workers' views of themselves as adults. The researchers chose for this purpose a **staged mixed-methods design**, in which a quantitative survey of 119 women workers, aged 18–29, from a factory in Guangdong, China, preceded and was the basis for qualitative interviews with 15 of them.

Zhong and Arnett posed three research questions:

1. To what extent do young Chinese migrant women workers consider themselves to have reached adulthood?
2. What are the conceptions of adulthood held by young Chinese migrant women workers—that is, what criteria are most important?
3. What is the relationship between marital status (or parental status) and conceptions of adulthood?

The researchers explained their use of mixed methods as reflecting the complexity of the phenomena involved with these questions:

> Given the complexity of social phenomena, mixed-methods research allows investigators to examine a problem from different but complementary perspectives, so that they can obtain a multilevel and contextual understanding of the phenomenon (Creswell and Plano Clark 2011). . . . In this study, explanatory design was implemented, in which qualitative findings were used to help interpret and contextualize quantitative results. (p. 257)

The structured questionnaire included questions about feelings of adulthood and the importance attached to various markers as indicating whether a person was an adult. Only 44% responded that they felt like they had reached adulthood, whereas most of the rest indicated that they felt adult in some ways but not in others. The five markers rated most often as important were (1) learn to care for parents; (2) if a man, become capable of supporting a family financially; (3) settle into a long-term career; (4) if a woman, become capable of caring for children; and (5) if a man, become capable of keeping family physically safe.

Analysis of the survey data indicated that being married and having children were associated with feeling like an adult, irrespective of age. The three markers of adulthood identified as most important in the quantitative survey were "learn to care for parents," "settle into a long-term career," and "become capable of caring for children."

The qualitative component began with the random selection of 15 of the women workers who had responded to the questionnaire.

> Two of the questions were asked to tap into their conceptions of adulthood. One was "Do you feel like you have reached adulthood? Why or why not?" which aimed to find out how the women workers viewed their transition to adulthood, and how they applied the markers of adulthood to their own life. The other question was about their more general conceptions of the transition to adulthood: "What do you think are the most important markers of adulthood?" Compared with the brief choice on the questionnaire, this question in the interview allowed the participants to describe their understanding of transition to adulthood in more detail and further explain which markers they thought were important for being an adult and why. (Zhong and Arnett 2014:258)

The qualitative analysis began with Zhong reading the transcripts and coding the themes identified about conceptions of adulthood. Arnett replicated this process, and the two compared their results and resolved discrepancies. Ultimately, they settled on four major domains that captured the expressed views about adulthood: family obligations and capacities, relational maturity, role transitions, and individualism. In the following sections of their article, each domain of markers for adulthood was elaborated with illustrations from the interviews.

As in the quantitative survey, the theme of family obligations and capacities was often emphasized as associated with feelings of adulthood. A quote from a married 23-year-old woman who had a one-year old son provided more insight into the basis for this association:

> I am married now, I should have a sense of responsibility toward family; I am also a mother, I should take care of my child. My parents-in-law are taking care of him, and we must send a few thousand yuan to them every year. Right now, he is only 1 year old. When it's time for him to go to school, I will go back home and take care of him myself. After all, my parents-in-law, who are not well educated, only care about if he is full [well-fed] or warm. They don't know how to teach him, so it's better for me to take care of him. (Zhong and Arnet 2014:260)

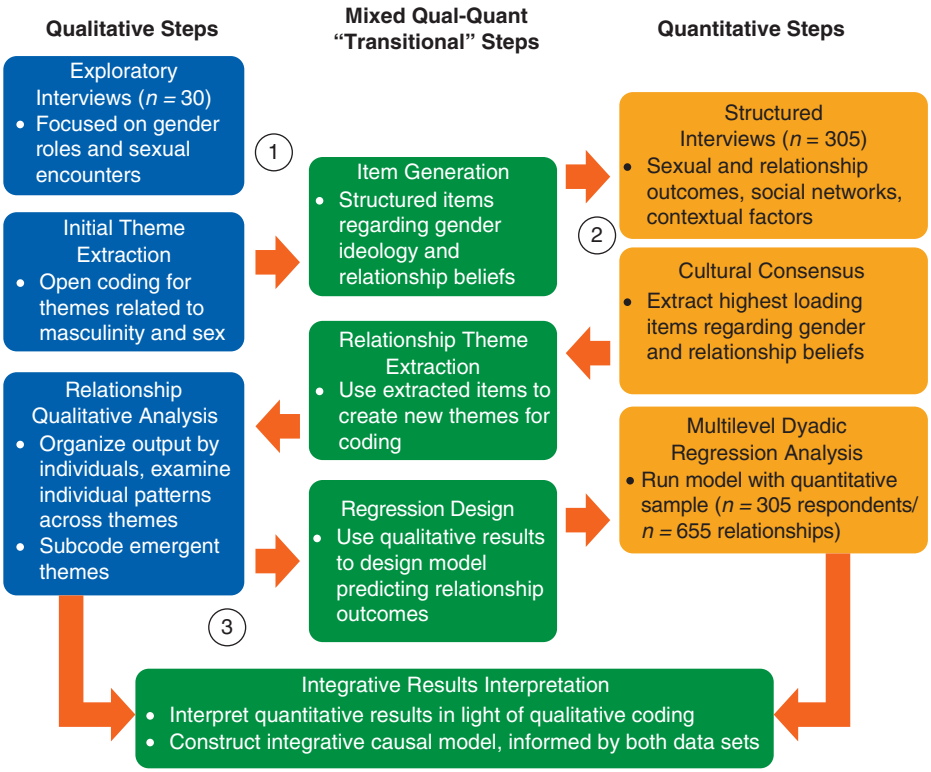
Complex Research Program Mixed-Methods Designs

Brown and his colleagues (2013) used a mixed-methods design in their study of sexual intimacy among homeless men in downtown Los Angeles. This study can best be understood as using a **complex research program mixed-methods design** because it involved a clear sequence in data collection as well as an integrated analysis in which both QUAL and QUAN methods had equal importance. The researchers first used qualitative interviews and then a structured quantitative survey, without giving either method priority. After they had collected data with both these QUAL and QUAN methods, Brown et al. developed an integrated analysis of the findings. The overall design might be diagrammed as a (QUAL→QUAN)→(QUAL+QUAN) design.

The qualitative research began with a sample of 30 persons using two meal lines and three shelters in the Los Angeles Skid Row area. Thirty of 36 men identified as eligible completed audiotaped interviews of 1 to 1 1/2 hours, for which they received $25. The researchers then finalized a structured questionnaire, based partly on what they learned in the qualitative interviews, and distributed this questionnaire several months later to a random sample of 338 men selected from 13 meal lines identified in service directories and provider interviews; 305 men completed the survey after giving their consent and then received a $25 incentive. The issues addressed in the qualitative interviews and structured survey overlapped in coverage of gender ideologies and sexual behaviors, and the structured survey also included questions about individual characteristics, sexual partners, and social network characteristics (Brown et al. 2013:329–330).

Figure 12.1 represents the different elements and stages of Brown et al.'s collection and analysis of data. In the first exploratory step, they conducted qualitative interviews and extracted themes from the resulting transcripts. These themes were then used in the second step to generate questions for their structured interviews. After they conducted the interviews, quantitative analysis of the structured data identified beliefs that were associated with each other and so suggested themes that could be examined with the qualitative data. Further analysis of the qualitative data then suggested more analyses with the quantitative data. Finally, the results of the qualitative and quantitative analyses were integrated into a combined explanation for the men's behavior.

FIGURE 12.1 ■ Qualitative and Quantitative Methodological Steps


Qualitative Steps
Mixed Qual-Quant "Transitional" Steps
Quantitative Steps
Exploratory Interviews (n = 30)
• Focused on gender roles and sexual encounters
1
Item Generation
• Structured items regarding gender ideology and relationship beliefs
Structured Interviews (n = 305)
• Sexual and relationship outcomes, social networks, contextual factors
2
Initial Theme Extraction
• Open coding for themes related to masculinity and sex
Cultural Consensus
• Extract highest loading items regarding gender and relationship beliefs
Relationship Theme Extraction
• Use extracted items to create new themes for coding
Relationship Qualitative Analysis
• Organize output by individuals, examine individual patterns across themes
• Subcode emergent themes
Multilevel Dyadic Regression Analysis
• Run model with quantitative sample (n = 305 respondents/ n = 655 relationships)
Regression Design
• Use qualitative results to design model predicting relationship outcomes
3
Integrative Results Interpretation
• Interpret quantitative results in light of qualitative coding
• Construct integrative causal model, informed by both data sets

Source: Brown, Ryan A., David P. Kennedy, Joan S. Tucker, Daniela Golinelli and Suzanne L. Wenzel. 2013. "Monogamy on the Street: A Mixed Methods Study of Homeless Men." *Journal of Mixed Methods Research* 7:3–28.

The analysis made possible by this integrated use of qualitative and quantitative data was actually even more complex than the figure indicates. In the authors' words:

> With interlaced qualitative and quantitative data collection and analysis, our initial gender-based model was subject to multiple revisions and modifications over time. Multiple rounds of qualitative and quantitative analysis (including multiple iterations of hypothesis generation and testing) moved us away from thinking about gender ideology and hypermasculinity and toward thinking about gender roles and structural barriers to enacting these roles. (Brown et al. 2013:330)

What emerged from the qualitative and quantitative elements of the analysis was an understanding of the homeless men as often upholding in theory the ideology of being responsible providers and partners but experiencing near constant frustration at being unable to live up to these widely shared ideals.

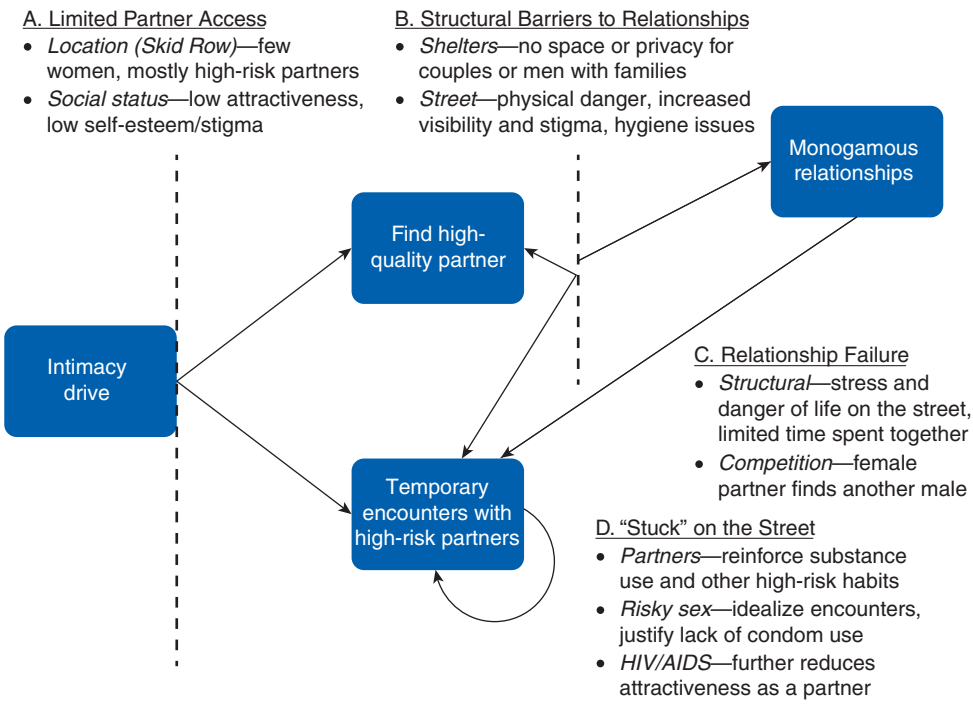
> There's something on my forehead that says homeless or not good enough or income bracket, you know what I mean? . . . Not so much what you're wearing, but where you're living. . . . They'll cut you off real quick. (Brown et al. 2013:336).

The quantitative data indicated the men were less likely to be monogamous if their partners were also homeless or abusing substances, but more likely to be monogamous if their partners were involved in the same social network (Brown et al. 2013:338).

The diagram in Figure 12.2 displays Brown et al.'s complete explanatory model of relationships.

> Overall, our analyses illustrated how men maintained hopes and dreams of idealized, committed relationships despite (for the most part) being unable to realize these dreams while living on the street. They also showed how these idealized hopes and dreams in some cases lead [sic] to or were used to justify risk behaviors. In particular, analyses pointed to a mixture of social and structural barriers to men actualizing a normatively desired (but infrequently attained), low-risk behavioral pathway—being in committed, monogamous relationships. (Brown et al. 2013:341)

FIGURE 12.2 ■ Holistic Model of High-Risk Sex Among Homeless Men


A. Limited Partner Access
• Location (Skid Row)—few women, mostly high-risk partners
• Social status—low attractiveness, low self-esteem/stigma
B. Structural Barriers to Relationships
• Shelters—no space or privacy for couples or men with families
• Street—physical danger, increased visibility and stigma, hygiene issues
Monogamous relationships
Find high-quality partner
Intimacy drive
C. Relationship Failure
• Structural—stress and danger of life on the street, limited time spent together
• Competition—female partner finds another male
Temporary encounters with high-risk partners
D. "Stuck" on the Street
• Partners—reinforce substance use and other high-risk habits
• Risky sex—idealize encounters, justify lack of condom use
• HIV/AIDS—further reduces attractiveness as a partner

Source: Brown, Ryan A., David P. Kennedy, Joan S. Tucker, Daniela Golinelli and Suzanne L. Wenzel. 2013. "Monogamy on the Street: A Mixed Methods Study of Homeless Men." *Journal of Mixed Methods Research* 7:3–28.

When different researchers studying the same research question over time have used different methods, some qualitative and some quantitative, we can also think of the series of investigations as an extended complexresearch program with mixed methods. The results of research using one method can in this way inform or raise questions about research focused on the same research question using a different research method. For example, Angela Moe's research (2007) on victim orientation to calling the police after an experience of abuse (Chapter 2's section on exploratory analysis) represented another step in the research program that began with Lawrence Sherman and Richard Berk's (1984) quantitative study of the police response to domestic violence. This represents a QUAN→QUAL research program, even though neither investigation was itself a mixed-methods study.

RESEARCH IN THE NEWS

CHILD ABUSE CASES DROP 51%. THE AUTHORITIES ARE VERY WORRIED.

New York City's child welfare agency received a weekly average of 672 child abuse or neglect cases in early spring 2020, compared to 1,374 in the same weeks in 2019. But this 51%t decline didn't mean kids were safer, Bronx district attorney Darcel Clark explained. "It's just the opposite." A methodological problem lay behind the apparent decline: The teachers, social workers, doctors and others who normally see kids regularly and file reports when they notice signs of abuse couldn't observe the children they knew during the lockdown. Some reports of abuse that did surface suggested awful circumstances for some children, as they were tied down or forced to care for themselves during the lockdown. Slowdowns in court procedures compounded problems. Agency personnel tried new approaches, visiting families and meeting with them outside instead of in their homes.

For Further Thought?

1. Propose a qualitative method and a quantitative method for investigating the likelihood of reporting instances of child abuse.
2. How could a mixed-methods design help determine whether there is a causal effect of child protection agency policies on the detection of cases of abuse and neglect?

News source: Stewart, Nikita. 2020. "Child Abuse Cases Drop 51 Percent. The Authorities Are Very Worried." *The New York Times*, June 9 (updated Aug. 7). (https://www.nytimes.com/2020/06/09/nyregion/coronavirus-nyc-child-abuse.html).

STRENGTHS AND LIMITATIONS OF MIXED METHODS

Combining qualitative and quantitative methods within one research project can strengthen the project's design by enhancing measurement validity, generalizability, causal validity, and/or authenticity. At the same time, combining methods creates challenges that may be difficult to overcome and ultimately limit the extent to which these goals are enhanced. I will illustrate these strengths and limitations with examples from studies introduced in this chapter and from the Boston McKinney Project, the major mixed-methods study of housing alternatives for homeless persons diagnosed with severe mental illness that was funded by the National Institute of Mental Health and the Department of Housing and Urban Development (Schutt 2011b).

Measurement validity is enhanced when questions to be used in a structured quantitative survey are first refined through qualitative cognitive interviewing or focus groups. After quantitative measures are used, qualitative measures can be added to clarify the meaning of numerical scores. You learned about this advantage of combining methods in Chapter 8, on survey research. Alternatively, quantitative measures can provide a more reliable indicator of the extent of variation between respondents that has already been described based on naturalistic observations or qualitative interviews. You learned in Chapter 10 how qualitative observations or interview material can be categorized and coded so that repeated instances can be counted. You will learn in Chapter 14 how online sources can provide complementary quantitative and qualitative data that can be analyzed with mixed-methods designs (Salmons 2016:124).

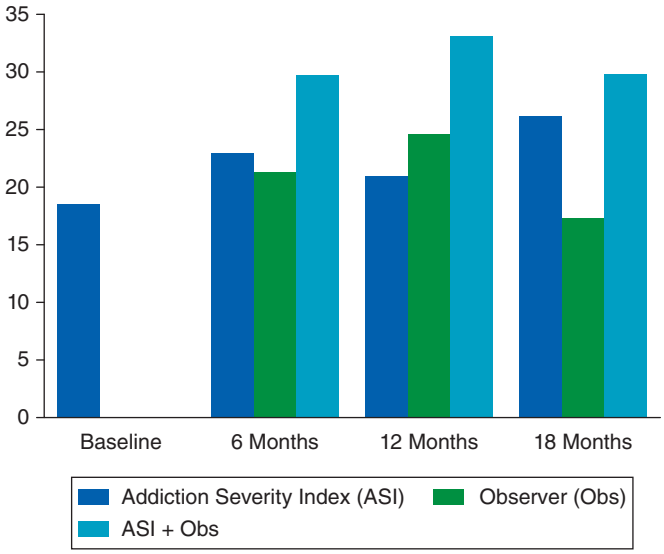
Measurement validity is also enhanced when measures using different methods result in a similar picture. If people are observed to do what they say they do, then the validity of a self-report measure is strengthened by measurement triangulation. The mixed-methods study of union democracy I introduced earlier in the chapter (Lipset et al. 1956; and see Schutt 1986)

achieved this goal in the assessment of the bases of union factions when the quantitative survey data led to a description of union factional divisions that was similar to that resulting from observations of union meetings and qualitative interviews with factional leaders. McCarter's (2009) qualitative interviews on juvenile justice decisions corroborated the impression that emerged from quantitative data of racial bias in decisions.

But what if different measures do not reinforce each other? Did some respondents forget what they have done, or are they purposely not disclosing their activities? Are interview respondents trying to "look good" in interviews by inflating their rate of participation in socially desirable activity or by minimizing their rate of participation in socially undesirable behavior (Brenner 2012)?

My use of observational data about substance use in group homes in addition to quantitative survey data in *Homelessness, Housing, and Mental Illness* (Schutt 2011b) provides an interesting example of how to use mixed methods to understand the bases of a partial failure of triangulation. Our quantitative measures included both self-report measures of substance abuse and observer (case manager) ratings of substance abuse using a structured form. As indicated in Figure 12.3, these different quantitative measures did not entirely agree. Some substance abusers did not report substance use that the case managers observed, and the case managers did not observe all the substance abuse problems that substance abusers reported. We termed those who did not report substance abuse during a period that they were observed to abuse substances as "nondisclosers" (Goldfinger et al. 1996:671).

FIGURE 12.3 ■ Self-Report and Observer Ratings of Substance Abuse, by Follow-Up



Source: Adapted from Goldfinger, Stephen M., Russell K. Schutt, Larry J. Seidman, Winston M. Turner, Walter E. Penk, and George S. Tolomiczenko. "Self-Report and Observer Measures of Substance Abuse Among Homeless Mentally Ill Persons in the Cross-Section and Over Time." *The Journal of Nervous and Mental Disease* 184(11):667–672. Copyright © 1996, Wolters Kluwer, Lippincott, Williams & Wilkins. Reprinted with permission.

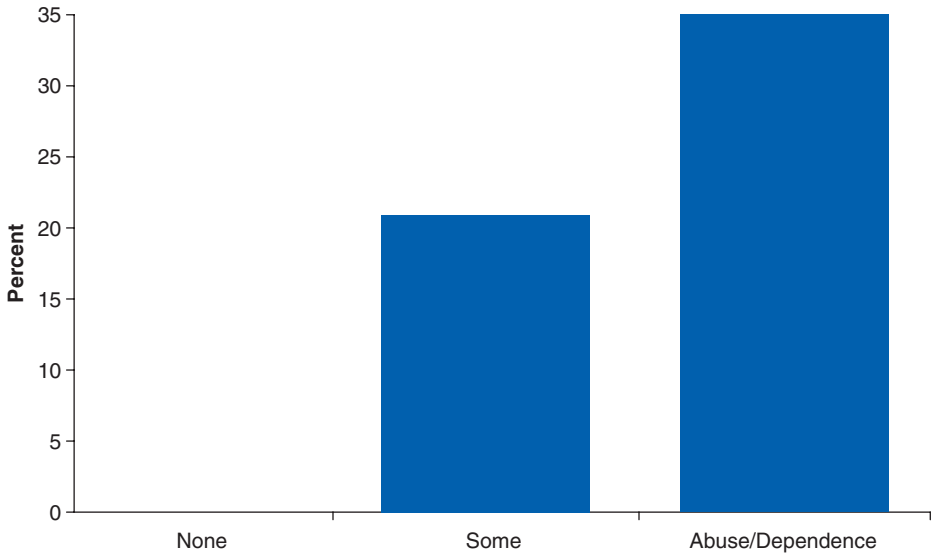
Observational notes recorded by project ethnographers helped explain the discrepant findings using these different measures. Some substance users would not acknowledge a problem that other project participants complained about:

> Tenants who were abusing alcohol and drugs often did not acknowledge any problem. One drinker explained that he "lost his toes [to frostbite] because he was on the street, not because he was drinking." Another argued that "the drinking cures mental illness" and another "says he is drinking because people aggravate him and don't give him the respect he deserves." . . . By contrast, tenants who did not have a substance abuse history or who were maintaining their sobriety often complained to staff that they felt threatened by others who were drinking. When one tenant insisted in a meeting that there was not a house drinking problem, another tenant retorted, "I'll disagree with that. Half the people in this house are drunk half the time." (Schutt 2011b:128)

In this example, qualitative data suggested that the failure of alternative quantitative measures to successfully triangulate resulted from the unwillingness of some of those who abused substances to acknowledge a drinking problem. The result of using mixed methods was thus a better understanding of the social reality underlying the individual behavior being measured. What at first might appear to be a limitation of mixed methods—the failure of alternative measures to converge, or triangulate—could instead be understood as an important insight resulting from the use of mixed methods.

The most common way that causal, or internal, validity is strengthened with a mixed-methods design is when qualitative interviews or observations are used to explore the mechanism involved in a causal effect. My mixed-methods analysis of the value of group and independent living options for people who had been homeless and diagnosed with severe mental illness demonstrates also how ethnographic data can help explain causal effects identified with an experimental design. Figure 12.4 displays the quantitative association between lifetime substance abuse—a diagnosis recorded on a numerical scale that was made on the basis of an

FIGURE 12.4 ■ Substance Abuse and Housing Loss in Group Homes



Source: HOMELESSNESS, HOUSING, AND MENTAL ILLNESS by Russell K. Schutt, r, Cambridge, Mass.: Harvard University Press, Copyright © 2011 by the President and Fellows of Harvard College. Used by permission. All rights reserved.

interview with a clinician—and housing loss—another quantitative indicator from service records (Schutt 2011b:135).

The ethnographic notes recorded in the group homes revealed orientations and processes that helped explain the pronounced association between substance abuse and housing loss (Schutt 2011b):

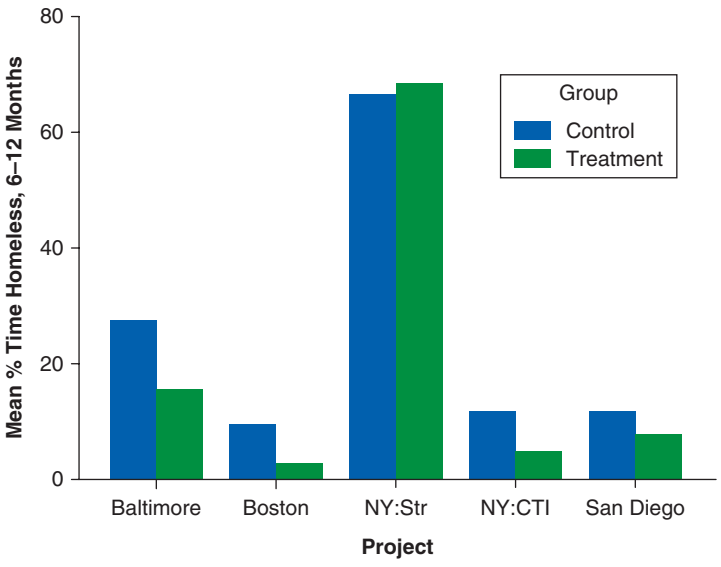
> The time has come where he has to decide once and for all to drink or not. . . . Tom has been feeling "pinned to the bed" in the morning. He has enjoyed getting high with Sammy and Ben, although the next day is always bad. . . . Since he came back from the hospital Lisandro has been acting like he is taunting them to throw him out by not complying with rules and continuing to drink. . . . (pp. 131, 133)

In this way, my analysis of the quantitative data reveals *what* happened, and my analysis of the ethnographic data helps to understand *why*. The same could be said for the way that qualitative interviews allowed Testa et al. (2011) to develop insights into how excessive alcohol consumption led to the quantitative association they identified between drinking and rape. It proved to be well worth the effort to collect qualitative data that would help interpret the results of the quantitative research.

A mixed-methods design can also improve external validity when a quantitative study is repeated in different contexts. Qualitative comparisons between these different contexts can then help make sense of the similarities and differences between outcomes across these contexts and thus help identify the conditions for the effects.

It was possible to make this type of qualitative comparison of quantitative study outcomes with the five experimental projects funded by the National Institute of Mental Health to identify the value of enhanced housing and services for reducing homelessness among homeless persons diagnosed with serious mental illness (our project in Boston was one of these five) (Schutt et

FIGURE 12.5 ■ Mean Total Time Homeless by Project



Source: Adapted from Russell K. Schutt, Richard L. Hough, Stephen M. Goldfinger, Antony F. Lehman, David I. Shern, Elie S. Valencia, and Patricia A. Wood. 2009. "Lessening Homelessness Among Persons With Mental Illness: A Comparison of Five Randomized Treatment Trials." *Asian Journal of Psychiatry* 2(3):100–105.

al. 2009). Figure 12.5 displays the quantitative comparison for the impact of enhanced housing (the treatment) between these five projects. The results show that enhanced housing reduced the time spent homeless in four of the five projects. The one project that was an exception to this rule was in New York City and involved a focus on moving homeless persons off the streets into a shelter, which is an improvement in residential status but still counts as being homeless (see Schutt 2011b:250–251). Interpretation of such multisite results must always be considered carefully, given the differences in project design that inevitably occur when the experiment is adapted to the specific conditions of each site.

The generalizability of qualitative findings can also be improved when a representative sample developed with a quantitative design is used to identify cases to study more intensively with qualitative methods. This was the approach used by Zhong and Arnett (2014) when they selected Chinese women for qualitative interviews from among those who had participated in their structured survey. More commonly, the combination of methods to improve generalizability occurs in a research program in which the two studies are conducted sequentially, most often by different researchers.

Mixed methods also facilitate achieving the goal of authenticity within a research project that also seeks to achieve measurement validity, generalizability, or causal validity. Brown et al.'s (2013) complex mixed-method study of homeless men leads to a conclusion that seems much more authentic than would have been the case if they had used just one method. Their analysis moved them away from simplistic assumptions about the men's conceptions of appropriate gender roles to a deeper understanding of their gender role aspirations in the context of limited means for achieving them. Tierolf et al.'s (2021) surprising finding that family violence did not increase among vulnerable families during the pandemic lockdown in the Netherlands was made more credible by their description of the patterns of adaptation they identified through the qualitative interviews.

It is naïve to think of mixed methods as simply overcoming problems that can occur when a research design is limited to one method. It is not always clear how best to compare qualitative and quantitative findings or how to interpret the discrepancies that arise after such comparison (Morgan 2014:66–81). These complexities, with unanticipated elements and insights, result in what some have called a certain "messiness" in mixed-methods research (Plano Clark and Ivankova 2016:276–277). We can't be certain that differences in findings mean that deficits have been uncovered in one method— as when some substance abusers do not disclose their abuse in answer to questions—or that the two methods are really answering different research questions—for example, what is the association between drinking and rape on campus, compared with how often does drinking result in incapacitation prior to a rape?

Mixed methods also create extra challenges for researchers because different types of expertise are required for the effective use of quantitative and qualitative methods. Recruiting multiple researchers for a project who then work as a team from conception to execution of the project may be the best way to overcome this limitation, but it may be difficult to mesh researchers with different backgrounds and perspectives (Plano Clark and Ivankova 2016:225, 228). The researchers also must acknowledge in planning the study timetable that the time required for collection, coding, and analysis of qualitative data can challenge a quantitative researcher's expectation of more rapid progress.

Despite these challenges, the different types of evidence produced in a mixed-methods investigation can strengthen overall confidence in the research findings and result in a more holistic understanding of the social world. Weaver-Hightower (2013) captured these advantages

in his methodological reflections on his mixed-method study of influences on public policy in Australia:

> Overall, this mixed-methods process, moving iteratively from qualitative to quantitative back to qualitative, established well the political and ideological influences on the policy makers and their "policy." The quantitative and qualitative methods were highly integrated—that is, the whole of the findings exceeded the sum of the individual quantitative and qualitative parts . . . because the quantitative procedures in some cases solidified, and in other cases challenged, my qualitative impressions of influence that were hard-won by hanging around, talking to people, and reading about the subject. Indeed, I was surprised by several influentials identified using the embedded quantitative phase. Several groups and individuals, like Canberra Grammar School or Boys in Focus, largely escaped my notice before the mixed methods approach. Likewise, without the qualitative case descriptions and the negative and positive case analyses, I would have been less able to understand why certain groups and actors were influential. And, regarding negative cases, the qualitative methods were still necessary to identify some influentials not netted by the mixed methods. In the end, then, the methods were integrated because both were necessary to find and evaluate influentials. (p. 17)

ETHICS AND MIXED METHODS

Researchers who combine methods must be aware of the ethical concerns involved in using each of the separate methods, but there are also some ethical challenges that are heightened in mixed-methods projects. One special challenge is defining the researcher's role in relation to the research participants. Every researcher creates an understanding about their role with research participants (Mertens 2012). Researchers using quantitative methods often define themselves as outside experts who design a research project and collect research data using objective procedures that are best carried out without involvement of stakeholders from the areas studied. By contrast, qualitative researchers often define themselves as engaging in research in some type of collaboration with the community or group they are studying, with much input from their stakeholders into the research design and the collection and analysis of research data.

A researcher using mixed methods cannot simply adopt one of these roles: A researcher needs some degree of autonomy when designing quantitative research plans, but a researcher will not be able to collect intensive qualitative data if participants do not accord the researcher some degree of trust as an insider. The challenge is compounded by the potential for different reactions of potential participants to the different roles. Authorities who control access to program clients or employees or to community members may be willing to agree to a structured survey but not to a long-term engagement with researchers as participant observers, so a mixed-methods project that spans programs, communities, or other settings may involve a biased sampling for the qualitative component. Luxardo et al. (2011) confronted this challenge in their study of Brazilian family violence services and as a result, focused their qualitative research on one service that supported the value of giving voice to their service recipients.

Weighing both roles and the best combination of them is critical at the outset of a mixed-methods project, although the dilemma will be lessened if a project uses different researchers to lead the quantitative and qualitative components. In our study of housing alternatives, for example, a team of ethnographers collected data on activities in the group homes while research assistants supervised by a different leader collected the project's quantitative data (Schutt 2011b).

Complex mixed-methods projects in which quantitative surveying is interspersed with observational research or intensive interviews may also require renegotiation of participant consent to the research procedures at each stage. As stated by Chih Hoong Sin (2005),

> Different stages and different components of research may require the negotiation of different types of consent, some of which may be more explicit than others. Sampling, contact, re-contact, and fieldwork can be underpinned by different conceptualization and operationalization of "informed consent." This behooves researchers to move away from the position of treating consent-seeking as an exercise that only occurs at certain points in the research process or only for certain types of research. Consent-seeking should not be thought of merely as an event. (p. 290)

In the qualitative component of their study of Brazilian victims of domestic violence, Luxardo and her colleagues (2011) adopted a flexible qualitative interviewing approach to allow participants to avoid topics they did not want to discuss:

> We tried to consider what was important for that adolescent during the interview and, many times, we had to reframe the content of the encounters according to the expectations they had. So, if they were not willing to share during an interview but still had complaints, doubts, or comments to share, we tried to focus on those instead of subtly directing the talk to the arena of the research interests. Moreover, we noticed that some adolescents (most of them migrants from Bolivia) did not feel at ease sharing that kind of information about their lives with a stranger, so we tried not to invade their intimacy by being culturally sensitive; if they did not want to talk, they did not have to do so. (p. 996)

CONCLUSIONS

A research project that is designed to answer multiple research questions and investigate a complex social setting often requires a mixed-methods design. Of course, to some extent, the complexity of the social world always exceeds what can be captured successfully with one method, but the challenges increase as our questions become more numerous and our social settings more complex. You have learned in this chapter about the different perspectives on combining methods, the different ways of mixing methods, and the challenges that arise when mixing methods. No matter what your methodological preference is at this point, increased understanding of these issues in mixed methods will improve your own research practice and your ability to critique the research of others.

I conclude with my justification for the mixed methods used in the project that was the foundation for my recent book about a very ambitious set of research questions in a complex and changing social setting (Schutt 2011b:284):

> *Homelessness, Housing, and Mental Illness* describes who participated in the Boston McKinney Project, what they wanted, and what they did; it evaluates whether living in group homes or independent apartments influenced participants' desires and actions; it explores how participants interacted with each other and whether they differed in their responses to the same stimuli. For these reasons and more, our project required systematic research methods—not just one method, such as conducting a survey, but multiple methods: different methods to answer different questions and to provide

> alternative perspectives on similar questions. . . . Even the best and most appropriate research methods do not entirely solve the problem of perspective. When we use social science methods, we see farther and probe deeper than we do in our everyday lives; however, no method—whether used alone or in combination with others—gives us perfect vision or infallible insight. Every method we use for investigating the social world will overlook some processes, distort some realities, and confuse some issues.

KEY TERMS

Complex mixed-methods design (p. 440)
Embedded mixed-methods design (p. 438)
Integrated mixed-methods design (p. 436)
Mixed methods(p. 432)
Multitrait–multimethod matrix (p. 432)
Paradigm wars (p. 434)
Staged mixed-methods design (p. 438)

HIGHLIGHTS

LO 12.1 Describe the development of mixed methods.

- Sociologists and other social scientists used mixed methods in community studies as long ago as the 1930s, but their popularity declined with the rapid growth of quantitative methods in the postwar period and during the subsequent period of the paradigm wars. Mixed methods have become much more popular in the past three decades.
- Multitrait–multimethod matrix: A method of evaluating the validity of measures by determining whether measuring the same phenomenon with different methods leads to convergent results and measuring different phenomena with the same method leads to divergent results.
- Triangulation: The use of multiple methods to study one research question; also used to mean the use of two or more different measures of the same variable.
- Mixed methods have been used more in applied fields than in academic disciplines.
- Paradigm wars: The intense debate from the 1970s to the 1990s between social scientists over the value of positivist and interpretivist/constructivist research philosophies and of using quantitative or qualitative methods; also see scientific paradigm.

LO 12.2 Diagram mixed methods that differ in sequence and priority.

- The primary method used in a mixed-methods project is written in all caps (QUAN or QUAL).
- The secondary method is written in lowercase letters (quan or qual).
- If both methods are given equal priority, they are both written in all caps.
- If one method is used before the other, the sequence is indicated with an arrow.
- If two methods are used concurrently but one has priority, the secondary method is said to be "embedded" in the primary method. This is indicated as follows: QUAL(quan) or QUAN(qual).
- If two methods are used concurrently but they have equal priority, the relation between the two methods is indicated with a +: QUAL+QUAN.

LO 12.3 Devise research studies that exemplify each major type of mixed methods.

- In an integrated mixed-methods design, qualitative and quantitative methods are used concurrently, and both are given equal priority.
- In an embedded mixed-methods design, qualitative and quantitative methods are used concurrently, but one has priority.
- In a staged mixed-methods design, one method is used in a sequence before the other, but they have equal priority.
- In a complex research program mixed-methods design, at least one of the four primary designs is combined with another of these designs, or one is used multiple times. This research program could involve one or more discrete research projects.

LO 12.4 Appraise the strengths and limitations of each major type of mixed method.

- Combining qualitative and quantitative methods within one research project can strengthen the project's design by enhancing measurement validity, generalizability, causal validity, and/or authenticity.
- Measurement validity is enhanced when questions to be used in a structured quantitative survey are first refined through qualitative cognitive interviewing or focus groups.
- Measurement validity is also enhanced when measures using different methods result in a similar picture.
- The most common way that causal, or internal, validity is strengthened with a mixed-methods design is when qualitative interviews or observations are used to explore the mechanism involved in a causal effect.
- A mixed-methods design can also improve external validity when a quantitative study is repeated in different contexts.
- Mixed methods also facilitate achieving the goal of authenticity within a research project that also seeks to achieve measurement validity, generalizability, or causal validity.
- However, it is naïve to think of mixed methods as simply overcoming problems that can occur when a research design is limited to one method.

- Mixed methods create extra challenges for researchers because different types of expertise are required for the effective use of quantitative and qualitative methods.
  - Despite the challenges, the different types of evidence produced in a mixed-methods investigation can strengthen overall confidence in the research findings and result in a more holistic understanding of the social world.

LO 12.5 Identify ethical issues in at least one example of a project using each type of mixed method.

- There are some ethical challenges that are heightened in mixed-methods projects.
- One special challenge is defining the researcher's role in relation to the research participants.
- Complex mixed-methods projects in which quantitative surveying is interspersed with observational research or intensive interviews may also require renegotiation of participant consent to the research procedures at each stage.

DISCUSSION QUESTIONS

1. Testa et al. (2011:237) claim, "Mixed-methods research can benefit any area of study," but also argue that this approach is particularly appropriate for the study of violence against women. What makes mixed methods more appropriate for research on the problem of violence against women, in their opinion? What about in your own opinion? Are there other areas that you think are more suited to the use of mixed methods? Explain your reasoning.
2. What other qualitative methods could be used to study violence against women on college campuses? What would be the advantages and disadvantages of focus groups or observation?
3. If you were asked to design a study of the impact of the pandemic on incidents of domestic violence and then develop a comprehensive report about the problem for the state government, would you put more emphasis on quantitative methods or qualitative methods, or would you attempt to use both equally? Compare your approach to that used by Tierolf et al. (2021) and explain why you adopted this approach.
4. Testa describes her own training as quantitative and then highlights some experiences that led her to integrate qualitative methods into her research (Testa et al. 2011). Would you describe your own training in research methods so far as primarily quantitative or qualitative? Have you had any experiences that led you to consider the "other" methodology?
5. Which of the four types of mixed methods do you feel is likely to be most useful for investigating the social world? Would you favor more single-method or more mixed-methods studies? Explain your reasoning.
6. Consider how ready you feel to design a mixed-methods research project. Do you think mixed-methods researchers should generally try to collaborate with another researcher who specializes in the methodology they are not so familiar with? Or should researchers seek to become experts in multiple methods so that they can combine them in research projects that they direct themselves?

SPSS EXERCISES

Could qualitative data help you understand quantitative associations between variables?

1. Generate the cross-tabulations of CONARMY_2, CONBUS_2, CONCLERG_2, CONEDU_2, CONFED_2, CONFINAN_2, CONJUDGE_2, CONLABOR_2, CONLEGIS_2, CONMEDIC_2 by EDUC4 in the GSS2020 data set (with column percents).
2. What do you learn about the relation between education and confidence in social institutions from these crosstabs? Be specific about differences between institutions.
3. What questions would you like to include in a qualitative interview with some GSS respondents to improve your understanding of the education–confidence association? Explain.

©iStockphoto.com/KatarzynaBialasiewicz

13 EVALUATION AND POLICY RESEARCH

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

13.1 Describe the history of evaluation research and its current status

13.2 Diagram the evaluation research process as a system

13.3 Define each of the types of evaluation research

13.4 Compare the alternatives in each of the five design decisions

13.5 Develop an argument for adding the policy research approach to evaluation research projects

13.6 Defend the ethics of research evaluations of social programs

Drug Abuse Resistance Education (D.A.R.E.) is offered in primary and secondary schools across the United States and in more than 50 other countries (D.A.R.E. 2021). Since its inception in 1983 as an innovative effort of the Los Angeles Police Department, D.A.R.E. has grown to include almost 1,700 programs and as many as one million elementary, middle, and high school students in the U.S. annually (Berman and Fox 2009; D.A.R.E. 2021). For parents worried about drug abuse among youth and for many concerned citizens, the program has immediate appeal. It brings a special police officer into the schools once a week to talk to students about the hazards of drug abuse and to establish a direct link between local law enforcement and young people. Although there are many substance abuse prevention programs, none have achieved the popularity or developed the infrastructure of D.A.R.E.

But does it work? Do students who participate in D.A.R.E. education become more resistant to the use of illicit drugs? Are they less likely to use illicit drugs while they are enrolled in the program or, more important, in the years after they have finished the program? Do students benefit in other ways from participation in D.A.R.E.? Are there beneficial effects for schools and communities? Although the idea of providing students with information about the harmful effects of substance abuse has intuitive appeal, the history of evaluation research about D.A.R.E. drives home an important point: To know whether social programs work and to understand how they work, we have to evaluate them systematically and fairly, whether we personally think the program sounds like a good idea or not.

This point was not lost on those charged with funding or administering D.A.R.E. The U.S. Department of Justice, which had funded the initial expansion of D.A.R.E. as a national program, also paid for a program evaluation by Susan Ennett, Christopher Ringwalt, and Robert Flewelling at the Research Triangle Institute in North Carolina, and Nancy Tobler at the State University of New York in Albany.Ennett and her colleagues (1994) located eight quantitative studies of the effects of D.A.R.E. that had used an experimental or quasi-experimental design with a focus on a particular state or locality. Across these studies, D.A.R.E. had no effect on drug or alcohol use at the time students completed D.A.R.E., although it led to a small reduction in tobacco use (and see Ringwalt et al. 1994; West and O'Neal 2004). D.A.R.E. participants did improve their knowledge about substance use as well as their social skills related to resisting substance abuse, attitudes toward the police, attitudes about drug use, and self-esteem, but these positive

attitudinal effects were less than those identified in evaluations of other types of substance abuse prevention programs. A six-year randomized field experiment of D.A.R.E.'s effectiveness in a sample of Illinois schools also found no long-term beneficial effects (Rosenbaum and Hanson 1998). As a result, some school districts stopped using D.A.R.E. (Rosenbaum 2007).

Yet this was not the end of the story. Federal officials convened a meeting of researchers and D.A.R.E. administrators to consider program changes and more research. As a result, the Robert Wood Johnson Foundation funded substance abuse researcher Zili Sloboda at the University of Akron to develop a new educational approach and a rigorous evaluation of its long-term impact (Berman and Fox 2009). Surprisingly, Sloboda and her colleagues (2009) found that this new program, Take Charge of Your Life, led to *increased* use of alcohol and cigarettes and no change in marijuana use. So D.A.R.E. administrators rejected that approach, adopted a different model already in use ("keepin' it REAL," or "KiR"), and retooled once again (Toppo 2002; West and O'Neal 2004).

> Gone is the old-style approach to prevention in which an officer stands behind a podium and lectures students in straight rows. New D.A.R.E. officers are trained as "coaches" to support kids who are using research-based refusal strategies in high-stakes peer-pressure environments. (D.A.R.E. 2008)

Of course, the "new D.A.R.E." was then the focus of evaluation research projects. One early quasi-experimental evaluation in 17 urban schools, funded by D.A.R.E. America, found no effect of the program on students' substance use (Vincus et al. 2010). Some researchers concluded that the program should simply be ended, while others have concluded that some communities have had good reasons for continuing to offer the program (Berman and Fox 2009; Birkeland et al. 2005). In 2017, Theodore L. Caputi at the University of Pennsylvania's Wharton School of Business and A. Thomas McLellan at the Treatment Research Institute in Philadelphia (Caputi and McLellan 2017) reviewed all 11 published studies reporting results of tests of KiR's effectiveness in reducing substance use. At the time, there had been no experimental studies of KiR programs offered through D.A.R.E., so they reviewed other programs that had tested the KiR program. The evidence didn't support the conclusion that KiR was effective in reducing drug use when offered to elementary school students, which was the focus of most D.A.R.E. KiR programs.

This may seem like a depressing way to begin a chapter on evaluation research. If, like me, you have a child who enjoyed D.A.R.E. or were yourself a D.A.R.E. student, you may know how appealing the program can be and how important its message is. But that should help to drive home the key point: Government agencies and other bodies that fund social programs must invest the necessary resources to evaluate their effectiveness, no matter how appealing they seem on their surface. Resources are too scarce to spend millions of dollars on a national program that does not achieve its intended goals. As we review more of the story of evaluation research on D.A.R.E. and other programs in this chapter, you will also learn that program evaluations can give us much more insight into the complexity of the social world. Evaluation findings can convince us that our preconceptions about human motivations or program processes need to be revised; they can identify differences in how different people react to the same program and how the same program may have different effects in different social contexts; and they can alert us to the importance of some program outcomes—both harmful and beneficial—that we may have overlooked. Moreover, the efforts of federal officials to encourage reexamination of

and change in the D.A.R.E. program demonstrate that program evaluations can have real impacts on public policy.

In this chapter, you will read about a variety of social program evaluations as I introduce the evaluation research process, illustrate the different types of evaluation research, highlight alternative approaches, and review ethical concerns. You will learn in this chapter that current major debates about social policies, such as health care, homelessness, domestic violence, and drug abuse, are each being informed by important major evaluation projects and that broad "policy research" projects can help define new directions for public action. You should finish the chapter with a much better understanding of how the methods of applied social research can help improve society.

HISTORY OF EVALUATION RESEARCH

Evaluation research is not a method of data collection, like survey research or experiments, nor is it a unique component of research designs, like sampling or measurement. Instead, evaluation research is social research that is conducted for a distinct purpose: to investigate social programs (e.g., substance abuse treatment programs, welfare programs, criminal justice programs, or employment and training programs). For each project, an evaluation researcher must select a research design and a method of data collection that are useful for answering the particular research questions posed and appropriate for the particular program investigated. Policy research, introduced later in the chapter, leads to recommendations based on the results of multiple projects.

The development of evaluation research as a major enterprise followed on the heels of the expansion of the federal government during the Great Depression and World War II. Large Depression-era government outlays for social programs stimulated interest in monitoring program output, and the military effort in World War II led to some of the necessary review and contracting procedures for sponsoring evaluation research. In the 1960s, criminal justice researchers began to use experiments to test the value of different policies (Orr 1999:24). New government social programs of the 1960s often came with evaluation requirements attached, and more than 100 contract research and development firms began operation in the United States between 1965 and 1975 (Dentler 2002; Rossi and Freeman 1989:34). The first evaluation research journal—*Evaluation Review*—was launched by SAGE in 1976 (Fox, Grimm, and Caldeira 2016:16).

The New Jersey Income Maintenance Experiment was the first large-scale randomized experiment to test social policy in action. Designed in 1967, the New Jersey experiment randomly assigned 1,300 families to different income support levels to test the impact of cash transfers to the working poor on their work effort. It was soon followed by even larger experiments to test other income maintenance questions, most notably the Seattle–Denver Income Maintenance Experiment (Orr 1999:24–26). The number of social experiments like this continued to increase in subsequent years (Greenberg, Shroder, and Onstott 1999:159). By 2001, there were more than 10,000 citations to what appeared to be randomized trials in social, psychological, educational, and criminological research (Petrosino et al. 2001). By 2017, the Campbell Collaboration research archive (www.campbellcollaboration.org), a project begun by evaluation researchers in 1999, contained 304 systematic reviews of social research evaluation results for sets of investigations ranging in focus from crime and justice and social welfare to education and international development.

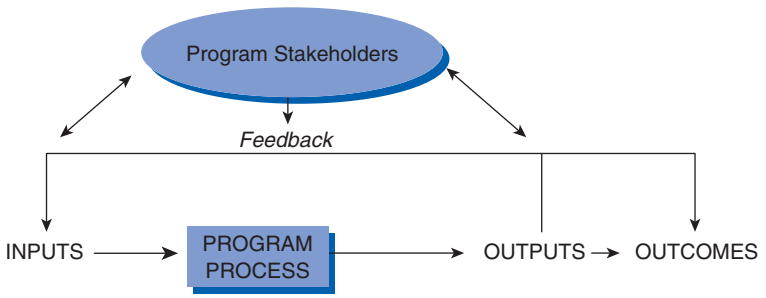
Government requirements and popular concern with the benefit derived from taxpayer-funded programs continue to stimulate evaluation research. The Community Mental

Health Act Amendments of 1975 (Public Law 94–63) required quality assurance reviews, which often involved evaluation-like activities (Patton 2002:147–151), and the Government Performance and Results Act of 1993 required some type of evaluation of all government programs (U.S. Office of Management and Budget 2002). In 1998, the federal Safe and Drug-Free Schools and Communities Act was reauthorized with the proviso that all programs to be funded had to be research based (Petrosino et al. 2001). At century's end, the federal government was spending about $200 million annually on evaluating $400 billion in domestic programs, and the 30 major federal agencies had between them 200 distinct evaluation units (Boruch 1997). In 1999, the new Governmental Accounting Standards Board urged that more attention be given to "service efforts and accomplishments" in standard government fiscal reports (Campbell 2002). This same period saw increasing requirements for performance indicators, performance audits, and accreditation reviews of public programs, which can themselves stimulate interest in formal evaluation research designed to improve performance (Fox et al. 2016:5–6). Nonprofit organizations like Britain's Social Care Institute for Excellence (www.scie.org.uk) encourage adoption of evidence-based practices.

EVALUATION BASICS

Figure 13.1 illustrates the process of evaluation research as a simple systems model. First, clients, customers, students, or some other persons or units—cases—enter the program as **inputs**. (You'll notice that this model treats programs like machines, with people functioning as raw materials to be processed.) Students may begin a new school program, welfare recipients may enroll in a new job training program, or crime victims may be sent to a victim advocate. The resources and staff a program requires are also program inputs.

FIGURE 13.1 ■ A Model of Evaluation



Source: Adapted from Martin and Kettner (1996).

Next, some service or treatment is provided to the cases. This may be attendance in a class, assistance with a health problem, residence in new housing, or receipt of special cash benefits. The **program process** may be simple or complicated, short or long, but it is designed to have some impact on the cases.

The direct product of the program's service delivery process is its **output**. Program outputs may include clients served, case managers trained, food parcels delivered, or arrests made. The program outputs may be desirable in themselves, but they primarily indicate that the program is operating.

Program **outcomes** indicate the impact of the program on the cases that have been processed. Outcomes can range from improved test scores or higher rates of job retention to fewer criminal offenses and lower rates of poverty. Any social program is likely to have multiple outcomes—some intended and some unintended, some positive and others that are viewed as negative.

Variation in both outputs and outcomes, in turn, influences the inputs to the program through a **feedback** process. If not enough clients are being served, recruitment of new clients may increase. If too many negative side effects result from a trial medication, the trial may be limited or terminated. If a program does not appear to lead to improved outcomes, clients may go elsewhere.

Evaluation research is simply a systematic approach to feedback: It strengthens the feedback loop through credible analyses of program operations and outcomes. Evaluation research also broadens this loop to include connections to parties outside the program itself. A funding agency or political authority may mandate the research, outside experts may be brought in to conduct the research, and the evaluation research findings may be released to the public, or at least the funders, in a formal report.

The evaluation process can be understood only in relation to the interests and perspectives of program stakeholders. **Stakeholders** are those individuals and groups who have some basis of concern with the program. They might be clients, staff, managers, funders, or the public. The board of a program or agency, the parents or spouses of clients, the foundations that award program grants, the auditors who monitor program spending, the members of Congress—each is a potential stakeholder, and each has an interest in the outcome of any program evaluation. Some may fund the evaluation; some may provide research data; and some may review or even approve the research report (Martin and Kettner 1996:3). Who the program stakeholders are and what role they play in the program evaluation will have tremendous consequences for the research.

Can you see the difference between evaluation research and traditional social science research (Posavac and Carey 1997)? Unlike explanatory social science research, evaluation research is not designed to test the implications of a social theory; instead, it is designed to answer a basic practical question, such as: What is the program's impact? Process evaluation often uses qualitative methods like traditional social science does, but unlike exploratory research, the goal is not to induce a broad theoretical explanation for what is discovered. Instead, the question is, how does the program do what it does? Unlikeresearch motivated only by academic concerns, evaluation researchers cannot design evaluation studies simply in accord with the highest scientific standards and the most important research questions; instead, the program stakeholders set the agenda. But there is no sharp boundary between the evaluation research and traditional social science research. In their attempt to explain how and why a program has an impact and whether the program is needed, evaluation researchers often bring social theories into their projects but for immediately practical aims.

RESEARCH IN THE NEWS
NEW STUDY ELUCIDATES THE BENEFITS OF PRE-K

Boston expanded its pre-K program in the public schools in the late 1990s, but the program did not have sufficient capacity to accept most students. The Boston Public Schools decided to use a lottery to decide which children were able to enroll. This lottery system allowed a randomized experiment to identify the benefits of pre-school for student outcomes. Three economists evaluated the results. Although the children who won the lottery did not have higher scores on standardized tests through high school, they did do better in many other

more concrete ways. They were less likely to be suspended in school or to be incarcerated as juveniles and, they were more likely to graduate from high school and to enroll in college. Children's social and emotional skills and overall maturity seemed to improve as a result of the pre-K experience, which seems far more important than improvements on standardized tests.

For Further Thought?

1. What methods would you propose to investigate the process by which the pre-K program had its effects on these outcomes?
2. How do these results suggest the importance of measuring multiple outcomes, and what other outcomes do you suggest be examined?

News source: Leonhardt, David. 2021. "New Study Elucidates the Benefits of Pre-K." *The New York Times*, May 10. Published online as "The Power of Pre-K."

QUESTIONS FOR EVALUATION RESEARCH

Evaluation projects can focus on several questions related to the operation of social programs and the impact they have:

- Is the program needed?
- Can the program be evaluated?
- How does the program operate?
- What is the program's impact?
- How efficient is the program?

The specific methods used in an evaluation research project depend partly on which of these foci the project has.

Needs Assessment

Is a new program needed or an old one still required? Is there a need at all? A **needs assessment** attempts to answer these questions with systematic, credible evidence. Need may be assessed by social indicators, such as the poverty rate or the level of home ownership; by interviews of local experts, such as school board members or team captains; by surveys of populations in need; or by focus groups composed of community residents (Rossi and Freeman 1989).

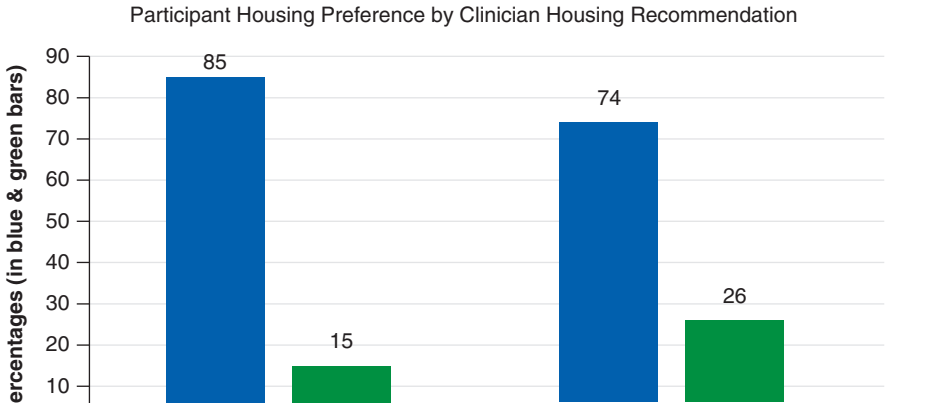
Needs assessment is not as easy as it sounds (Posavac and Carey 1997). Whose definitions or perceptions should be used to shape our description of the level of need? How will we deal with unawareness of need? How can we understand the level of need without understanding the social context from which that level of need emerges? (Short answer to that one: We can't!) What, after all, does *need* mean in the abstract? We won't really understand what the level of need is until we develop plans for implementing a program in response to the identified needs.

The results of the Boston McKinney Project reveal the importance of taking a multidimensional approach to the investigation of need. The Boston McKinney Project evaluated the merits of providing formerly homeless mentally ill persons with staffed group housing rather than with individual housing (Schutt 2011b). In a sense, you can think of the whole experiment as

involving an attempt to answer the question "What type of housing do these persons 'need'?" Our research team first examined this question at the start of the project by asking each project participant which type of housing they wanted (Schutt and Goldfinger 1996) and by independently asking two clinicians to estimate which of the two housing alternatives would be best for each participant (Goldfinger and Schutt 1996).

Figure 13.2 displays the findings. The clinicians recommended staffed group housing for 69% of the total 108 participants, whereas 78% of the participants themselves preferred individual housing. There was almost no correspondence between the housing recommendations of the clinicians and the housing preferences of the participants (who did not know what the clinicians had recommended for them). So which perspective reveals the level of need for staffed group housing versus individual housing?

FIGURE 13.2 ■ Type of Residence: Preferred and Recommended



Source: Based on Schutt 2011b:87 (Figure 4.9).

Yet another perspective on housing needs is introduced by the project's outcomes. Individuals assigned to the group housing were somewhat more successful in retaining their housing than were those who were assigned to individual housing, and this differential success rate grew in the years after the project's end (Schutt 2011b:247). Does this therefore reveal that these homeless mentally ill persons "needed" group housing more than they needed individual housing, despite their preference? What should we make of the fact that the participants who preferred individual housing were more likely to lose their housing during the project, whereas the participants whom the clinicians had rated as ready for independent living were less likely to lose their housing (Schutt 2011b:248)? And what should we make of the fact that whether or not participants received the type of housing the clinicians recommended or that they themselves preferred made no difference to the likelihood of their losing their housing during the project (Schutt 2011b:248)? Does this mean that neither initial preferences nor clinician recommendations tell us about the need for one or the other type of housing, only about the risk of losing whatever housing they were assigned to?

The methodological lesson here is that in needs assessment, as in other forms of evaluation research, it is a good idea to use multiple indicators. You can also see that there is no absolute definition of *need* in this situation, nor is there likely to be in any but the most simplistic evaluation projects. A good evaluation researcher will do their best to capture different perspectives on need and then help others make sense of the results.

A wonderful little tale, popular with evaluation researchers, reveals the importance of thinking creatively about what people need:

> The manager of a 20-story office building had received many complaints about the slowness of the elevators. He hired an engineering consultant to propose a solution. The consultant measured traffic flow and elevator features and proposed replacing the old with new ones, which could shave 20 seconds off the average waiting time. The only problem: It cost $100,000. A second consultant proposed adding 2 additional elevators, for a total wait time reduction of 35 seconds and a cost of $150,000. Neither alternative was affordable. A third consultant was brought in. He looked around for a few days and announced that the problem was not really the waiting times, but boredom. For a cost of less than $1,000, the manager had large mirrors installed next to the elevators so people could primp and observe themselves while waiting for an elevator. The result: no more complaints. Problem solved. (Witkin and Altschuld 1995:38)

Evaluability Assessment

Evaluation research will be pointless if the program itself cannot be evaluated. Yes, some type of study is always possible, but a study specifically to identify the effects of a particular program may not be possible within the available time and resources. So researchers may conduct an **evaluability assessment** to learn this in advance rather than expend time and effort on a fruitless project.

Why might a social program not be evaluable? Several factors may be involved:

- Management only wants to have its superior performance confirmed and does not really care whether the program is having its intended effects. This is a very common problem.
- Staff are so alienated from the agency that they don't trust any attempt sponsored by management to check on their performance.
- Program personnel are just "helping people" or "putting in time" without any clear sense of what the program is trying to achieve.
- The program is not clearly distinct from other services delivered from the agency and so can't be evaluated by itself (Patton 2002:164).

Because evaluability assessments are preliminary studies to "check things out," they often rely on qualitative methods. Program managers and key staff may be interviewed in depth, or program sponsors may be asked about the importance they attach to different goals. The evaluators may then suggest changes in program operations to improve evaluability. They may also use the evaluability assessment to "sell" the evaluation to participants and sensitize them to the importance of clarifying their goals and objectives. If the program is judged to be evaluable, knowledge gleaned through the evaluability assessment is used to shape evaluation plans.

Complex community initiatives can be particularly difficult to evaluate due to the evolving nature of the intervention as it is adapted to community conditions, an often broad range of outcomes, and—many times—the absence of a control or comparison community (Fox et al. 2016:43). The President's Family Justice Center (FJC) Initiative was such an initiative begun by President George W. Bush to plan and implement comprehensive domestic violence services that would provide "one-stop shopping" for victims in need of services. In 2004, the National Institute of Justice contracted with Abt Associates in Cambridge, Massachusetts, to assess the

evaluability of 15 pilot service programs that had been awarded a total of $20 million and to develop an evaluation plan. In June 2005, Abt researchers Meg Townsend, Dana Hunt, Caity Baxter, and Peter Finn reported on their evaluability assessment.

Abt's assessment began with conversations to collect background information and perceptions of program goals and objectives from those who had designed the program. These conversations were followed by a review of the grant applications submitted by each of the 15 sites and phone conversations with site representatives. Site-specific data collection focused on the project's history at the site, its stage of implementation, staffing plans and target population, program activities and stability, goals identified by the site's director, apparent contradictions between goals and activities, and the state of data systems that could be used in the evaluation. Table 13.1 shows the resulting logic model that illustrates the intended activities, outcomes, and impacts for the Alameda County, California, program. Although they had been able to begin the evaluability assessment process, Townsend and her colleagues concluded that in the summer of 2005, none of the 15 sites were far enough along with their programs to complete the assessment.

TABLE 13.1 ■ Alameda Family Justice Center Logic Model

Inputs	Activities	Outcomes	Impacts	Goals
• On-site partners • Intake systems • Client management process • Space design • Site location	**FJC** • Case management • Assistance with restraining orders • Assistance with police reports • Legal assistance • Advocacy • Medical care • Forensic exams • Assessments and referral for treatment • Counseling • Safety planning • Emergency food/cash/ transportation • Referral for shelter and other ongoing care • Assistance with public assistance • 24-hour helpline • Parenting classes • Child care • Rape crises services • Faith-based services • Job training • Translation services	**Victims** • Increase likelihood to access services • Increase demand for services • Increase usage of services • Increase frequency of crossreferrals or use of multiple services	**Victims** • Reduce tendency to blame oneself for abuse • Reduce conditions that prevent women from leaving • Increase likelihood of reporting incident • Increase likelihood of request for temporary/ permanent restraining orders • Increase likelihood of participating in prosecution	• Decrease incidents of DV* • Decreased repeat victimizations • Decreased seriousness • Hold offenders accountable • Decrease repeat offenders • Break cycle of violence

Inputs	Activities	Outcomes	Impacts	Goals
	Community ● Early intervention and prevention programming ● FJC informational materials	**Community** ● Increase knowledge of DV/SA*/elder abuse ● Increase awareness of services available	**Community** ● Increase awareness of FJC ● Decrease social tolerance for VAW*	
	Systems ● Collaboration between government and nongovernment providers ● Improve access to batterer information	**Systems** ● Improve DV policies and procedures ● Increase understanding of each other's services ● Increase coordination of services	**Systems** ● Improve institutional response to DV ● Decrease secondary trauma ● Increase assurance of victim safety ● Increase the number of successful criminal legal actions ● Increase the number of successful civil legal actions	

*DV = domestic violence; SA = substance abuse; VAW = violence against women.

Source: Meg Townsend, Dana Hunt, Caity Baxter, and Peter Finn. 2005. *Interim Report: Evaluability Assessment of the President's Family Justice Center Initiative.* Cambridge, MA: Abt Associates Inc. Reprinted with permission.

Process Evaluation

What actually happens in a social program? The New Jersey Income Maintenance Experiment was designed to test the effect of some welfare recipients receiving higher payments than others (Kershaw and Fair 1976). Did that occur? In the Minneapolis experiment on the police response to domestic violence (Sherman and Berk 1984), police officers were to either arrest or warn individuals accused of assaulting their spouses on the basis of a random selection protocol, unless they concluded that they must override the experimental assignment to minimize the risk of repeat harm. Did the police officers follow this protocol? How often did they override it because of concerns about risk? Questions like these about program implementation must be answered before it is possible to determine whether the program's key elements had the desired effect. Answers to such program implementation questions are obtained through **process evaluation**—research to investigate the process of service delivery.

Process evaluation is even more important when more complex programs are evaluated. Many social programs comprise multiple elements and are delivered over an extended period, often by different providers in different areas. Because of this complexity, it is quite possible that the program as delivered is not the same for all program recipients or consistent with the formal program design.

The evaluation of D.A.R.E. by Research Triangle Institute researchers Ringwalt and colleagues (1994:7) included a process evaluation with three objectives:

1. To assess the organizational structure and operation of representative D.A.R.E. programs nationwide
2. To review and assess the factors that contribute to the effective implementation of D.A.R.E. programs nationwide
3. To assess how D.A.R.E. and other school-based drug prevention programs are tailored to meet the needs of specific populations

The process evaluation (they called it an *implementation assessment*) was an ambitious research project in itself, with site visits, informal interviews, discussions, and surveys of D.A.R.E. program coordinators and advisers. These data indicated that D.A.R.E. was operating as designed and was running relatively smoothly. As shown in Table 13.2, drug prevention coordinators in D.A.R.E. school districts rated the program components as more satisfactory than did coordinators in school districts with other types of alcohol and drug prevention programs.

TABLE 13.2 ■ Components of D.A.R.E. and Other Alcohol and Drug Prevention Programs Rated as Very Satisfactory (%)

Component	D.A.R.E. Program (*N* = 222)	Other Programs (*N* = 406)
Curriculum	67.5%	34.2%
Teaching	69.7%	29.8%
Administrative requirements	55.7%	23.1%
Receptivity of students	76.5%	34.6%
Effects on students	63.2%	22.8%

Source: Ringwalt et al. (1994:58).

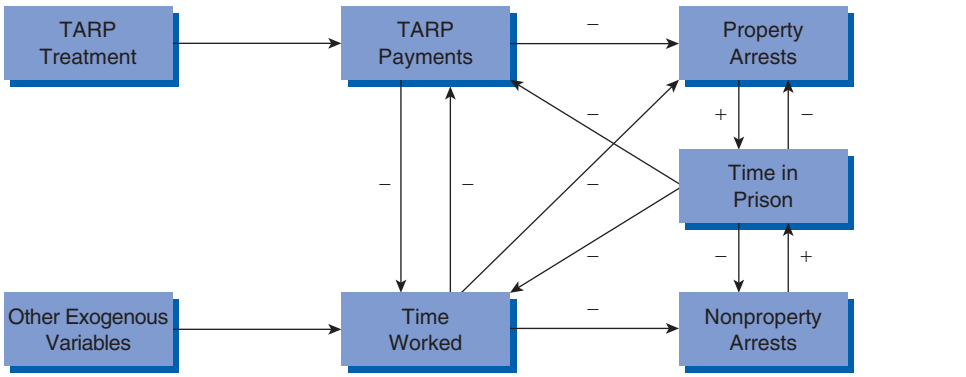
Process evaluation also can be used to identify the specific aspects of the service delivery process that have an impact. This, in turn, will help explain why the program has an effect and which conditions are required for these effects. (In Chapter 6, I described this as identifying the causal mechanism.) For instance, implementation problems identified in D.A.R.E. site visits included insufficient numbers of officers to carry out the program as planned and a lack of Spanish-language D.A.R.E. books in a largely Hispanic school. Classroom observations indicated engaging presentations and active student participation (Ringwalt et al. 1994:58). Sloboda et al.'s (2009) evaluation of the trial *Take Care of Your Life* curriculum found that D.A.R.E. officers delivered all of the lessons and 73% of the planned content in those lessons—a rate of program delivery that is better than that in many large-scale efforts (Sloboda et al. 2009:8).

Process analysis of this sort can show how apparent findings may be incorrect. The apparently disappointing results of the Transitional Aid Research Project (TARP) provide an instructive lesson of this sort. TARP was a social experiment designed to determine whether financial aid during the transition from prison to the community would help released prisoners find employment and avoid returning to crime. Two thousand participants in Georgia and Texas were randomized to receive either a particular level of benefits over a particular period or no benefits at all (the control group). Initially, it seemed that the payments had no effect: The rate

of subsequent arrests for both property and nonproperty crimes was not affected by the TARP treatment condition.

But this wasn't all there was to it. Peter Rossi tested a more elaborate causal model of TARP effects, summarized in Figure 13.3 (Chen 1990). Participants who received TARP payments therefore had more income to begin with and so had more to lose if they were arrested; therefore, they were less likely to commit crimes. However, TARP payments also created a disincentive to work and therefore increased the time available in which to commit crimes. Thus, the positive direct effect of TARP (more to lose) was cancelled out by its negative indirect effect (more free time).

FIGURE 13.3 ■ Model of TARP Effects



Source: Chen (1990:210). Reprinted with permission from SAGE Publications, Inc.

The term **formative evaluation** may be used instead of process evaluation when the evaluation findings are used to help shape and refine the program (Rossi and Freeman 1989). Formative evaluation procedures that are incorporated into the initial development of the service program can specify the treatment process and lead to changes in recruitment procedures, program delivery, or measurement tools (Patton 2002:220). They can provide a strong foundation for managers as they develop the program (Fox et al. 2016:10)

You can see the formative element in the following excerpt from the report on Ohio's Assisted Living Medicaid Waiver Program by a research team from the Scripps Gerontology Center and Miami University (Applebaum et al. 2007). The program is designed to allow disabled elderly persons to move into assisted living facilities that maximize autonomy rather than more expensive nursing homes.

> In order for the state to develop a viable assisted living program, a plan to increase provider participation is critical. . . . The perspectives of residents, and their families, providers, case managers, and representatives of the Department of Aging and Health will be needed to refine the program as it develops. . . . Every program needs to evolve and therefore a solid structure to identify and implement necessary changes will be crucial for long-term program success. (p. 31)

Process evaluation can employ a wide range of indicators. Program coverage can be monitored through program records, participant surveys, community surveys, or users versus

dropouts and ineligibles. Service delivery can be monitored through service records completed by program staff, a management information system maintained by program administrators, or reports by program recipients (Rossi and Freeman 1989).

Qualitative methods are often a key component of process evaluation studies because they can be used to elucidate and understand internal program dynamics—even those that were not anticipated (Patton 2002:159; Posavac and Carey 1997). Qualitative researchers may develop detailed descriptions of how program participants engage with each other, how the program experience varies for different people, and how the program changes and evolves over time. The goal is to develop a "bottom-up" view of the process rather than the "top-down" view that emerges from official program records (Fox et al. 2016:64–66).

Impact Analysis

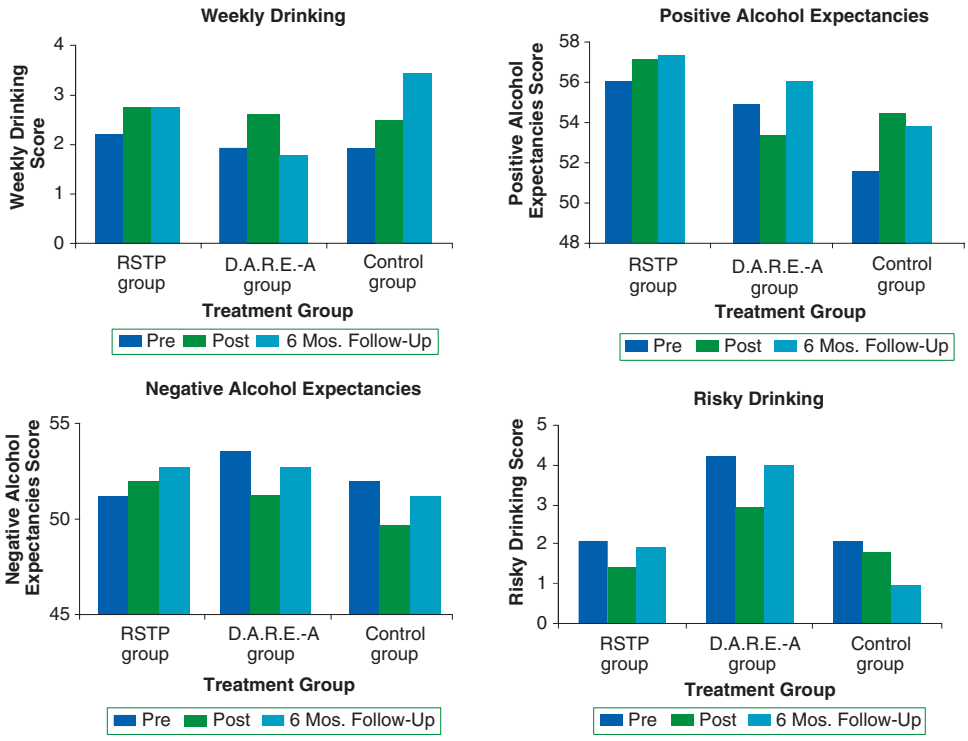
The core questions of evaluation research are "Did the program work?" and "Did it have the intended result?" This part of the research is variously called **impact analysis**, **impact evaluation**, or **summative evaluation**. Formally speaking, impact analysis compares what happened after a program with what would have happened had there been no program.

Think of the program—a new strategy for combating domestic violence, an income supplement, whatever—as manipulating an independent variable and the result it seeks as a change in a dependent variable. The D.A.R.E. program (independent variable), for instance, tries to reduce drug use (dependent variable). When the program is present, we expect less drug use. In a more elaborate study, we might have multiple values of the independent variable; for instance, we might look at *no program, D.A.R.E. program*, and *other drug/alcohol education* conditions and compare the results of each. In a complex intervention, it is important to measure program components carefully in order to determine whether specific components were responsible for any impacts identified (Fox et al. 2016:84).

Elizabeth J. D'Amico and Kim Fromme's (2002) study of a new Risk Skills Training Program (RSTP) is a good example of a more elaborate study. They compared the impact of RSTP on children aged 14 to 19 years with that of an abbreviated version of D.A.R.E. and with results for a control group. The impacts they examined included positive and negative alcohol expectancies (the anticipated effects of drinking) as well as perception of peer risk taking and actual alcohol consumption. They found that negative alcohol expectancies increased for the RSTP group in the posttest but not for the D.A.R.E. group or the control group. Moreover, weekly drinking and positive expectancies for drinking outcomes actually *increased* for the D.A.R.E. group or the control group by the six-month follow-up but not for the RSTP group (pp. 568–570; see Figure 13.4).

As in other areas of research, an experimental design is the preferred method for maximizing internal validity—that is, for making sure causal claims about program impact are justified. Cases are assigned randomly to one or more experimental treatment groups and to a control group so there is no systematic difference between the groups at the outset (see Chapter 7). The goal is to achieve a fair, unbiased test of the program itself so that differences between the types of people who are in the different groups do not influence judgment about the program's impact. It can be a difficult goal to achieve because the usual practice in social programs is to let people decide for themselves whether they want to enter a program or not and to establish eligibility criteria that ensure people who enter the program are different from those who do not enter the program (Boruch 1997). If people who begin the program differ from those who do not begin the program, a selection bias is introduced.

FIGURE 13.4 ■ Impact of RSTP, D.A.R.E.-A



Source: Based on D'Amico and Fromme (2002:569).

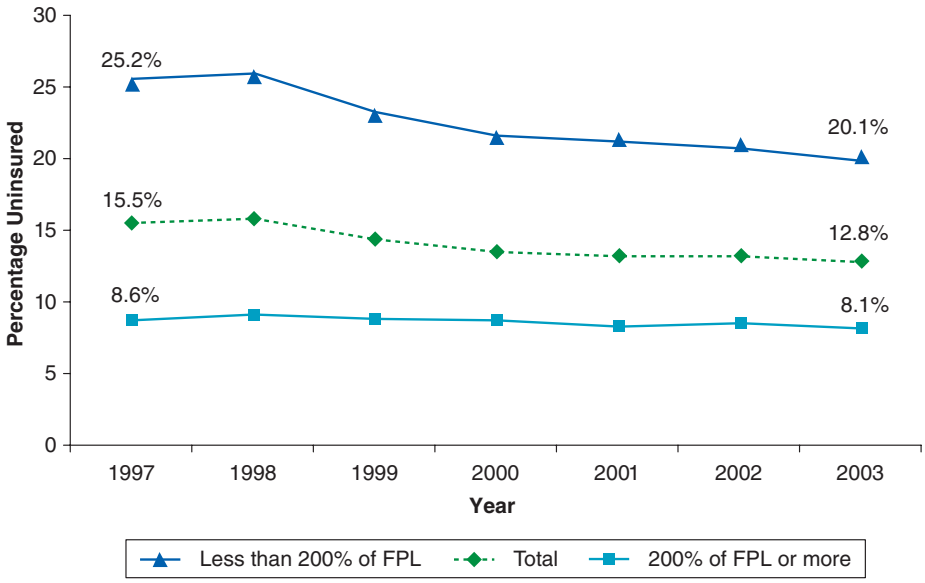
Impact analyses that do not use an experimental design can still provide useful information and may be all that is affordable, conceptually feasible, or ethically permissible in many circumstances. Evaluation of the State Children's Health Insurance Program (SCHIP) provides an example. The U.S. Congress enacted SCHIP in 1997 to expand health insurance coverage for low-income uninsured children. The federal Centers for Medicare and Medicaid Services (CMS) then contracted with Mathematica Policy Research to evaluate SCHIP (Rosenbach et al. 2007:1).

Given the nature of SCHIP, children could not be assigned randomly to participate. Instead, the Mathematica researchers tracked the growth of enrollment in SCHIP and changes in the percentage of children without health insurance. You can see in Figure 13.5 that the percentage of uninsured children declined from 25.2% to 20.1% from 1997 to 2003, during which time enrollment in SCHIP grew from 0.7 million to 6.2 million children (Rosenbach et al. 2007:ES.2). SCHIP, Margo Rosenbach and her colleaguesconcluded,

> provided a safety net for children whose families lost employer-sponsored coverage during the economic downturn [of 2000–2003] . . . [even as] nonelderly adults experienced a significant 2 percentage point increase in their overall uninsured rate. . . . In the absence of SCHIP . . . the number of uninsured children would have grown by 2.7 million, rather than declining by 0.4 million. (p. ES.7)

Of course, program impact may also be evaluated with quasi-experimental designs (see Chapter 7) or survey or field research methods. But if current participants who are already in

FIGURE 13.5 ■ Percentage of Children Under Age 19 Without Health Insurance by Poverty Level



Source: Margo Rosenbach, Carol Irvin, Angela Merrill, Shanna Shulman, John Czajka, Christopher Trenholm, Susan Williams, So Sasigant Limpa-Amara, and Anna Katz. 2007. *National Evaluation of the State Children's Health Insurance Program: A Decade of Expanding Coverage and Improving Access: Final Report.* Cambridge, MA: Mathematica Policy Research, Inc.

a program are compared with nonparticipants, it is unlikely that the treatment group will be comparable with the control group. Participants will probably be a selected group, different at the outset from nonparticipants. As a result, causal conclusions about program impact will be on much shakier ground. For instance, when a study at New York's maximum-security prison for women concluded "Income Education [i.e., classes] Is Found to Lower Risk of New Arrest," the findings were immediately suspect: The research design did not ensure that the women who enrolled in the prison classes were similar to those who had not enrolled in the classes, "leaving open the possibility that the results were due, at least in part, to self-selection, with the women most motivated to avoid re-incarceration being the ones who took the college classes" (Lewin 2001a:A18).

Rigorous evaluations often lead to the conclusion that a program does not have the desired effect (Patton 2002:154). A program depends on political support for achieving its goals, and such evaluations may result in efforts to redesign the program (as with D.A.R.E.) or reduction or termination of program funding. The latter outcome occurred with the largest U.S. federal training program for the disadvantaged, the Job Training Partnership Act (JTPA). In 1995, partly because of negative results of evaluation research, Congress both restructured JTPA and cut its funding by more than 80% (Heckman, Hohmann, and Smith 2000:651).

But the JTPA experience also raises an important caution about impact analysis. The National JTPA Study used a rigorous experimental design to evaluate the JTPA program and concluded that it did not have the desired impact. However, a subsequent, more intensive analysis by the economist James Heckman and his colleagues (2000) indicated that the initial evaluation had overlooked two critical issues: (1) between 27% and 40% of control group members found some other training program to participate in while the research was being

conducted, and (2) between 49% and 59% of treatment group members dropped out of the program and so did not actually receive the classroom training (p. 660). After Heckman et al. (2000) accounted for these problems of control group "substitution bias" and treatment group dropout, they concluded that classroom training in JTPA had "a large positive effect on monthly earnings" for those completing training (p. 688). There is even more of a basis for concern when program participants know that a special intervention is being tested and not everyone is receiving it. Staff delivering a program in the control condition could adopt some elements of the intervention being tested to try to help their clients, while program clients could become demoralized because they are not receiving the intervention (Fox et al. 2016:93). So it is important to design research carefully and consider all possible influences on program impact before concluding that a program should be terminated because of poor results.

Efficiency Analysis

Whatever the program's benefits, are they sufficient to offset the program's costs? Are the taxpayers getting their money's worth? What resources are required by the program? These efficiency questions can be the primary reason why funders require evaluation of the programs they fund. As a result, **efficiency analysis**, which compares program effects with costs, is often a necessary component of an evaluation research project.

A **cost–benefit analysis** must identify the specific program costs and the procedures for estimating the economic value of specific program benefits. This type of analysis also requires that the analyst identify whose perspective will be used to determine what can be considered a benefit rather than a cost. Program clients will have a different perspective on these issues than will taxpayers or program staff. Table 13.3 lists the factors that can be considered as costs or benefits in an employment and training program, from the standpoint of program participants, the rest of society, and society as a whole (the combination of program participants and the rest of society) (Orr 1999:224). Note that some anticipated impacts of the program, on welfare benefits and wage subsidies, are considered a cost to one group and a benefit to another group, whereas some are not relevant to one of the groups.

A **cost–effectiveness analysis** focuses attention directly on the program's outcomes rather than on the economic value of those outcomes. In a cost–effectiveness analysis, the specific costs of the program are compared with the program's outcomes, such as the number of jobs obtained, the extent of improvement in reading scores, or the degree of decline in crimes committed. For example, one result might be an estimate of how much it cost the program for each job obtained by a program participant.

Social science training often doesn't devote much attention to cost–benefit analysis, so it can be helpful to review possible costs and benefits with an economist or business school professor or student. Once the potential costs and benefits have been identified, they must be measured. This is a need highlighted in new government programs (Campbell 2002):

> The Governmental Accounting Standards Board's (GASB) mission is to establish and improve standards of accounting and financial reporting for state and local governments in the United States. In June 1999, the GASB issued a major revision to current reporting requirements ("Statement 34"). The new reporting will provide information that citizens and other users can utilize to gain an understanding of the financial position and cost of programs for a government and a descriptive management's discussion and analysis to assist in understanding a government's financial results. (p. 1)

TABLE 13.3 ■ Conceptual Framework for Cost–Benefit Analysis of an Employment and Training Program

Costs/Benefits	Perspective of Program Participants	Perspective of Rest of Society	Perspective of Entire Society*
Costs			
Operational costs of the program	0	–	–
Forgone leisure and home production	–	0	–
Benefits			
Earnings gains	+	0	+
Reduced costs of nonexperimental services	0	+	+
Transfers			
Reduced welfare benefits	–	+	0
Wage subsidies	+	–	0
Net benefits	±	±	±

*Entire society = program participants + rest of society.

– = program costs; + = program benefits; ± = program costs and benefits; 0 = no program costs or benefits.

Source: Orr, Larry L. 1999. *Social Experiments: Evaluating Public Programs With Experimental Methods.* Thousand Oaks, CA: Sage.

In addition to measuring services and their associated costs, a cost–benefit analysis must be able to make some type of estimation of how clients benefited from the program. Normally, this will involve a comparison of some indicators of client status before and after clients received program services or between clients who received program services and a comparable group that did not.

A recent study of therapeutic communities (TCs) provides a clear illustration. A TC is a method for treating substance abuse in which abusers participate in an intensive, structured living experience with other addicts who are attempting to stay sober. Because the treatment involves residential support as well as other types of services, it can be quite costly. Are those costs worth it?

Stanley Sacks and colleagues (2002) conducted a cost–benefit analysis of a modified TC. In the study, 342 homeless, mentally ill chemical abusers were randomly assigned to either a TC or a "treatment-as-usual" comparison group. Employment status, criminal activity, and utilization of health care services were each measured for the 3 months before entering treatment and the 3 months after treatment. Earnings from employment in each period were adjusted for costs incurred by criminal activity and utilization of health care services.

Was it worth it? The average cost of TC treatment for a client was $20,361. In comparison, the economic benefit (based on earnings) to the average TC client was $305,273, which declined to $273,698 after comparing postprogram with preprogram earnings, and it was still $253,337, even after adjustment for costs. The resulting benefit–cost ratio was 13:1, although this ratio

declined to only 5.2:1 after further adjustments (for cases with extreme values). Nonetheless, the TC program studied seems to have had a substantial benefit relative to its costs.

DESIGN DECISIONS

Once we have decided on or identified the goal or focus of a program evaluation, there are still important decisions to be made about how to design the specific evaluation project. The most important decisions are the following:

- *Black box evaluation or program theory:* Do we care how the program gets results?
- *Researcher or stakeholder orientation:* Whose goals matter the most?
- *Quantitative or qualitative methods:* Which methods provide the best answers?
- *Simple or complex outcomes:* How complicated should the findings be?
- *Groups or individuals:* Assign groups or individuals to different programs?

Black Box Evaluation or Program Theory

The "meat and potatoes" of most evaluation research involves determining whether a program has the intended effect. If the effect occurred, the program has "worked"; if the effect didn't occur, then, some would say, the program should be abandoned or redesigned. In **black box evaluation**, the process by which a program has an effect on outcomes is treated as a black box—that is, the focus of the evaluation researcher is on whether cases seem to have changed as a result of their exposure to the program between the time they entered the program as inputs and when they exited the program as outputs (Chen 1990). The assumption is that program evaluation requires only the test of a simple input–output model, like that shown in Figure 13.1. There may be no attempt to open the black box of the program process.

But there is good reason to open the black box and investigate how the process works (or why it doesn't work). Consider recent research on welfare-to-work programs. The Manpower Demonstration Research Corporation reviewed findings from research on these programs in Florida, Minnesota, and Canada. In each location, adolescents with parents in a welfare-to-work program were compared with a control group of teenagers whose parents were on welfare but were not enrolled in welfare-to-work. In all three locations, teenagers in the welfare-to-work families actually did worse in school than those in the control group—troubling findings.

But why? Why did requiring welfare mothers to work hurt their children's schoolwork? Unfortunately, because the researchers had not investigated the program process—had not "opened the black box"—we can't know for sure. Martha Zaslow, an author of the resulting research report, speculated,

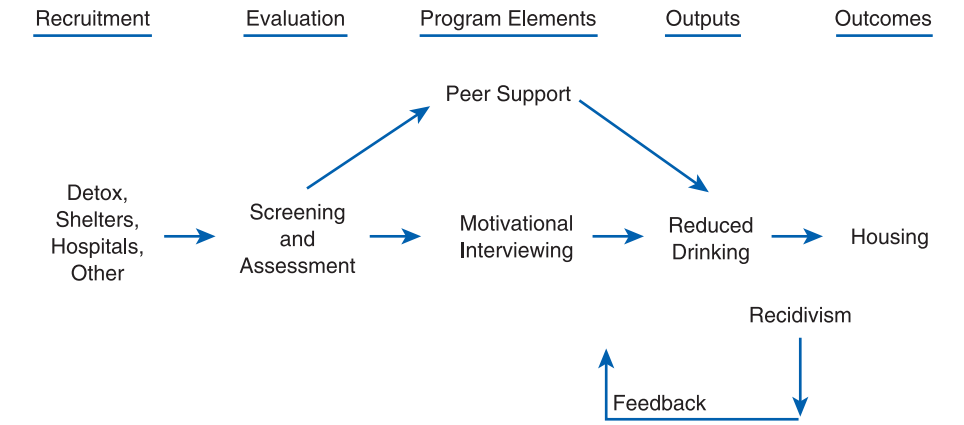
> Parents in the programs might have less time and energy to monitor their adolescents' behavior once they were employed. . . . Under the stress of working, they might adopt harsher parenting styles . . . the adolescents' assuming more responsibilities at home when parents got jobs was creating too great a burden. (cited in Lewin 2001b:A16)

But as Zaslow admitted, "We don't know exactly what's causing these effects, so it's really hard to say, at this point, what will be the long-term effects on these kids" (cited in Lewin 2001b:A16).

If an investigation of program process is conducted, a **program theory** may be developed. A program theory specifies how the program is expected to operate and identifies which program elements are operational (Chen 1990:32). It may identify the resources needed by the program and the activities that are essential to its operations, as well as how the program is to produce its effects (Fox et al. 2016:44). A program theory thus improves understanding of the relationship between the independent variable (the program) and the dependent variable (the outcome or outcomes). When a researcher has sufficient knowledge of this before the investigation begins, outlining a program theory can help guide the investigation of a program process in the most productive directions. This is termed a **theory-driven evaluation**.

For example, Figure 13.6 illustrates the theory for an alcoholism treatment program. It shows that persons entering the program are expected to respond to the combination of motivational interviewing and peer support. A program theory can also decrease the risk of failure when the program is transported to other settings because it will help identify the conditions required for the program to have its intended effect.

FIGURE 13.6 ■ The Program Theory for a Treatment Program for Homeless Persons With Alcohol Addictions



When it is possible to conduct multiple evaluation experiments, a program theory can be tested by repeating the impact analysis with different elements of the program omitted (Fox et al. 2016:96–97). For example, you can imagine repeating the impact analysis represented in Figure 13.6 after removing the peer support element of the treatment program and then after removing instead the motivational interviewing element of the program.

Program theory can be either descriptive or prescriptive (Chen 1990). Descriptive theory specifies what impacts are generated and how they occur. It suggests a causal mechanism, including intervening factors, and the necessary context for the effects. Descriptive theories are generally empirically based. Conversely, prescriptive theory specifies what the program ought to do but is not actually tested. Prescriptive theory specifies how to design or implement the treatment, what outcomes should be expected, and how performance should be judged. Comparison of the descriptive and prescriptive theories of the program can help identify implementation difficulties and incorrect understandings that can be corrected (Patton 2002:162–164).

Researcher or Stakeholder Orientation

Whose prescriptions specify how the program should operate, what outcomes it should try to achieve, or who it should serve? Most social science research assumes the researcher specifies the research questions, the applicable theory or theories, and the outcomes to be investigated. Social science research results are most often reported in a professional journal or at professional conferences, where scientific standards determine how the research is received. In program evaluation, however, the research question is often set by the program sponsors or the government agency responsible for reviewing the program. In consulting projects for businesses, the client—a manager, perhaps, or a division president—decides what question researchers will study. Research findings are reported to these authorities. Most often, these authorities also specify the outcomes to be investigated. The first evaluator of the evaluation research is the funding agency, then, rather than the professional social science community. Evaluation research is research for a client, and its results may directly affect the services, treatments, or even punishments (e.g., in the case of prison studies) that program users receive. In this case, the person who pays the piper gets to call the tune.

Should the evaluation researcher insist on designing the evaluation project and specifying its goals, or should they accept the suggestions and adopt the goals of the funding agency? What role should the preferences of program staff or clients play? What responsibility does the evaluation researcher have to politicians and taxpayers when evaluating government-funded programs? The different answers that various evaluation researchers have given to these questions are reflected in different approaches to evaluation (Chen 1990:66–68).

Social science (or researcher) approaches emphasize the importance of researcher expertise and maintenance of some autonomy to develop the most trustworthy, unbiased program evaluation. It is assumed that "evaluators cannot passively accept the values and views of the other stakeholders" (Chen 1990:78). Evaluators who adopt this approach derive a program theory from the information they obtain on how the program operates and extant social science theory and knowledge, not from the views of stakeholders. In one somewhat extreme form of this approach, *goal-free evaluation*, researchers do not even permit themselves to learn what goals the program stakeholders have for the program. Instead, the researcher assesses and then compares the needs of participants with a wide array of program outcomes (Scriven 1972b). The goal-free evaluator wants to see the unanticipated outcomes and to remove any biases caused by knowing the program goals in advance.

Stakeholder approaches encourage researchers to be responsive to program stakeholders (so this approach is also termed **responsive evaluation**). Issues for study are to be based on the views of people involved with the program, and reports are to be made to program participants (Shadish, Cook, and Leviton 1991:275–276). The program theory is developed by the researcher to clarify and develop the key stakeholders' theory of the program (Shadish et al. 1991:254–255). In one stakeholder approach, termed *utilization-focused evaluation*, the evaluator forms a task force of program stakeholders, who help to shape the evaluation project so they are most likely to use its results (Patton 2002:171–175). In evaluation research, termed *action research* or *participatory research* (discussed in Chapter 11), program participants are engaged with the researchers as coresearchers and help to design, conduct, and report the research. One research approach that has been termed *appreciative inquiry* eliminates the professional researcher altogether in favor of a structured dialogue about needed changes among program participants themselves (Patton 2002:177–185).

In their book *Fourth Generation Evaluation*, Egon Guba and Yvonna Lincoln (1989) argue for evaluations oriented toward stakeholders:

> The stakeholders and others who may be drawn into the evaluation are welcomed as equal partners in every aspect of design, implementation, interpretation, and resulting action of an evaluation—that is, they are accorded a full measure of political parity and control . . . determining what questions are to be asked and what information is to be collected on the basis of stakeholder inputs. (p. 11)

Because different stakeholders may differ in their reports about or assessment of the program, there is not likely to be one conclusion about program impact. The evaluators are primarily concerned with helping participants understand the views of other stakeholders and with generating productive dialogue. Tineke Abma (2005) took this approach in a study of an injury prevention program at a dance school in the Netherlands:

> The evaluators acted as facilitators, paying deliberate attention to the development of trust and a respectful, open and comfortable climate. . . . Furthermore, the evaluation stimulated a public discourse about issues that were taboo, created a space for reflection, fostered dynamics and motivated participants to think about ways to improve the quality of their teaching practice. (pp. 284–285)

Of course, there are disadvantages in both stakeholder and social science approaches to program evaluation. If stakeholders are ignored, researchers may find that participants are uncooperative, that their reports are unused, and that the next project remains unfunded. However, if social science procedures are neglected, standards of evidence will be compromised, conclusions about program effects will likely be invalid, and results are unlikely to be generalizable to other settings. These equally undesirable possibilities have led to several attempts to develop more integrated approaches to evaluation research.

Integrative approaches attempt to cover issues of concern to both stakeholders and evaluators and to include stakeholders in the group from which guidance is routinely sought (Chen and Rossi 1987:101–102). The emphasis given to either stakeholder or social science concerns is expected to vary with the specific project circumstances. Integrated approaches seek to balance the goal of carrying out a project that is responsive to stakeholder concerns with the goal of objective, scientifically trustworthy, and generalizable results. When the research is planned, evaluators are expected to communicate and negotiate regularly with key stakeholders and to consider stakeholder concerns. Findings from preliminary inquiries are reported back to program decision makers, so they can make improvements in the program before it is formally evaluated. When the actual evaluation is conducted, the evaluation research team is expected to operate more autonomously, minimizing intrusions from program stakeholders.

Many evaluation researchers now recognize that they must account for multiple values in their research and be sensitive to the perspectives of different stakeholders, in addition to maintaining a commitment to the goals of measurement validity, internal validity, and generalizability (Chen 1990). Ultimately, evaluation research takes place in a political context, in which program stakeholders may be competing or collaborating to increase program funding or to emphasize particular program goals. A political process creates social programs, and a political process determines whether these programs are evaluated and what is done with the evaluation findings (Weiss 1993:94). Developing supportive relations with stakeholder groups will increase the odds that political processes will not undermine evaluation practice. You don't want to find out after you are finished that "people operating ineffective programs who depend on them for

their jobs" are able to prevent an evaluation report from having any impact ("'Get Tough' Youth Programs" 2004:25).

Quantitative or Qualitative Methods

Evaluation research that attempts to identify the effects of a social program typically is quantitative: Did the response times of emergency personnel tend to decrease? Did the students' test scores increase? Did housing retention improve? Did substance abuse decline? It's fair to say that when there's an interest in comparing outcomes between an experimental and a control group or tracking change over time in a systematic manner, quantitative methods are favored.

But qualitative methods can add much to quantitative evaluation research studies, including more depth, detail, nuance, and exemplary case studies (Patton 2002). Perhaps the greatest contribution qualitative methods can make in many evaluation studies is investigating program processes—finding out what is inside the black box. Although it is possible to track service delivery with quantitative measures such as frequency of staff contact and number of complaints, finding out what is happening to clients and how clients experience the program can often best be accomplished by observing program activities and interviewing staff and clients intensively.

For example, Michael Quinn Patton (2002:160) describes a study in which process analysis in an evaluation of a prenatal clinic's outreach program led to program changes. The process analysis revealed that the outreach workers were spending a lot of time responding to immediate problems, such as needs for rat control, protection from violence, and access to English classes. As a result, the outreach workers were recruiting fewer community residents for the prenatal clinic. New training and recruitment strategies were adopted to lessen this deviation from program goals.

Another good reason for using qualitative methods in evaluation research is the importance of learning how different individuals react to the treatment. For example, a quantitative evaluation of student reactions to an adult basic skills program for new immigrants relied heavily on the students' initial statements of their goals. However, qualitative interviews revealed that most new immigrants lacked sufficient experience in the United States to set meaningful goals; their initial goal statements simply reflected their eagerness to agree with their counselors' suggestions (Patton 2002:177–181).

Qualitative methods can also help reveal how social programs actually operate. Complex social programs have many different features, and it is not always clear whether the combination of those features or some particular features are responsible for the program's effect—or for the absence of an effect. Lisbeth Schorr, director of the Harvard Project on Effective Interventions, and Daniel Yankelovich, president of Public Agenda, put it this way: "Social programs are sprawling efforts with multiple components requiring constant midcourse corrections, the involvement of committed human beings, and flexible adaptation to local circumstances" (Schorr and Yankelovich 2000:A19).

The more complex the social program is, the more value qualitative methods can add to the evaluation process. Schorr and Yankelovich (2000) discuss the Ten Point Coalition, an alliance of black ministers that helped reduce gang warfare in Boston through multiple initiatives, "ranging from neighborhood probation patrols to safe havens for recreation" (p. A19). Qualitative methods would help describe a complex, multifaceted program such as this. A skilled qualitative researcher will be flexible and creative in choosing methods for program evaluation and will often develop mixed methods (described in Chapter 12), so the evaluation benefits from the advantages of both qualitative and quantitative techniques.

Simple or Complex Outcomes

Does the program have only one outcome? Unlikely. How many outcomes are anticipated? How many might be unintended? Which are direct consequences of program action, and which are indirect effects that occur as a result of the direct effects (Mohr 1992)? Do the longer-term outcomes follow directly from the immediate program outputs? Does the output (e.g., the increase in test scores at the end of the preparation course) result surely in the desired outcomes (e.g., increased rates of college admission)? Because of these and other possibilities, the selection of outcome measures is a critical step in evaluation research.

The decision to focus on one outcome rather than another, on a single outcome, or on several can have enormous implications. When Lawrence Sherman and Richard Berk (1984) evaluated the impact of an immediate arrest policy in cases of domestic violence in Minneapolis, they focused on recidivism as the key outcome. Similarly, the reduction of recidivism was the single desired outcome of prison "boot camps" opened in the 1990s. Boot camps are military-style programs for incarcerated persons that provide tough, highly regimented activities and harsh punishment for disciplinary infractions, with the goal of scaring the incarcerated person "straight." Boot camps were quite the rage in the 1990s, and the researchers who evaluated their impact understandably focused on criminal recidivism.

But these single-purpose programs turned out to be not quite so simple to evaluate. The Minneapolis researchers found there was no adequate single source for records of recidivism in domestic violence cases, so they had to hunt for evidence from court and police records, follow-up interviews with victims, and family member reports (Sherman and Berk 1984). More easily measured variables, such as partners' ratings of the accused's subsequent behavior, eventually received more attention. Boot camp researchers soon concluded the experience did not reduce recidivism: "Many communities are wasting a great deal of money on those types of programs" (Robert L. Johnson, cited in "'Get Tough' Youth Programs", 2004:25). However, some participants felt the study had missed something (Latour 2002):

> [A staff member] saw things unfold that he had never witnessed among inmates and their caretakers. Those experiences profoundly affected the drill instructors and their charges, who still call to talk to the guards they once saw as torturers. Graduation ceremonies routinely reduced inmates, relatives, and sometimes even supervisors to tears. (p. B7)

A former boot camp superintendent, Michael Corsini, compared the Massachusetts boot camp with other correctional facilities and concluded, "Here, it was a totally different experience" (Latour 2002:B7).

Some observers now argue the failure of boot camps to reduce recidivism results from the lack of postprison support rather than failure of the camps to promote positive change in incarcerated persons. Looking only at recidivism rates would be to ignore some important positive results.

Despite the additional difficulties introduced by measuring multiple outcomes, most evaluation researchers attempt to do so (Mohr 1992). When a program is evaluated initially only as a demonstration project, it is particularly important to consider the range of effects that might occur when it is adopted more widely in different communities (Fox et al. 2016:85). The result usually is a much more realistic and richer understanding of program impact.

Evaluation research on D.A.R.E. has also examined multiple outcomes. Most often, program impact is distinguished for type of drug use: alcohol, marijuana, and tobacco (Ennett et al. 1994:1399). Attitudinal change is also often examined. One common positive finding is of

improved police–community relations and a more positive image of law enforcement in the eyes of students (Sloboda et al. 2009:7).

> It's a very positive program for kids . . . a way for law enforcement to interact with children in a nonthreatening fashion . . . D.A.R.E. sponsored a basketball game. The middle school jazz band played. . . . We had families there. . . . D.A.R.E. officers lead activities at the [middle school]. . . . Kids do woodworking and produce a play. (Taylor 1999:1, 11)

For some people, this impact justifies maintaining the program even despite negative findings about its impact on drug abuse (Birkeland et al. 2005:248).

Measuring multiple outcomes may also lead to identification of different program impacts for different groups. In their evaluation of the alternative Take Charge of Your Life (TCYL) program for D.A.R.E. schools, for example, Sloboda and her colleagues (2009) found beneficial effects for students who reported marijuana use at baseline but harmful effects on alcohol and drug use for white students and concluded the most effective approach to prevention may differ for those who have already had experience with illicit substances.

Project New Hope was an ambitious experimental evaluation of the impact of guaranteeing jobs to poor persons (DeParle 1999). It was designed to answer the following question: If low-income adults are given a job at a sufficient wage, above the poverty level, with child care and health care assured, how many would ultimately prosper? Some of the multiple outcomes measured in the evaluation of Project New Hope appear in Table 13.4.

TABLE 13.4 ■ Outcomes in Project New Hope

Income and Employment (2nd program year)	**New Hope**	**Control Group**
Earnings	$6,602	$6,129
Wage subsidies	$1,477	$862
Welfare income	$1,716	$1,690
Food stamp income	$1,418	$1,242
Total income	$11,213	$9,915
% above poverty level	27%	19%
% continuously unemployed for 2 years	6%	13%
Hardships and Stress	**New Hope**	**Control Group**
% reporting:		
Unmet medical needs	17%	23%
Unmet dental needs	27%	34%
Periods without health insurance	49%	61%
Living in overcrowded conditions	14%	15%
Stressed much or all of the time	45%	50%
Satisfied or very satisfied with standard of living	65%	67%

Sources: DeParle (1999); Bos et al. (1999).

The project involved 677 low-income adults in Milwaukee who were offered a job involving work for 30 hours a week, along with child care and health care benefits. The outcome? Only 27% stuck with the job long enough to lift themselves out of poverty, and their earnings as a whole were only slightly higher than those of a control group that did not receive guaranteed jobs. Levels of depression were not decreased, and levels of self-esteem were not increased by the job guarantee. But there were some positive effects: The number of people who never worked at all declined, and rates of health insurance and use of formal child care increased. Perhaps most important, the classroom performance and educational hopes of participants' male children increased, with the boys' test scores rising by the equivalent of 100 points on the SAT and their teachers ranking them as better behaved.

So did the New Hope program "work"? Clearly it didn't live up to initial expectations, but it certainly showed that social interventions can have some benefits. Would the boys' gains continue through adolescence? Longer-term outcomes would be needed. Why didn't girls (who were already performing better than the boys) benefit from their parents' enrollment in New Hope just as the boys did? A process analysis would have added a great deal to the evaluation design. The long and short of it is that the collection of multiple outcomes gave a better picture of the program's impact.

Of course, there is a potential downside to the collection of multiple outcomes. Policy makers may choose to publicize only those outcomes that support their own policy preferences and ignore the rest. Often, evaluation researchers themselves have little ability to publicize a more complete story.

Groups or Individuals

Robert St. Pierre and Peter Rossi (2006) urge evaluation researchers to consider randomizing groups rather than individuals to alternative programs in an impact analysis. For example, a study of the effectiveness of a new educational program could randomly assign classes to either the new program or an alternative program rather than randomly assign individual students. Sloboda and her colleagues (2009) used this approach in their study of the alternative D.A.R.E. program, randomly assigning school districts to either the TCYL program or the control condition.

Randomization of groups to different treatments may be preferable when the goal is to compare alternative programs so that different groups each receive some type of program. It can be easier to implement this approach if there are already different programs available and parents or other constituents are concerned they (or their children) receive *some* type of program. Using group randomization also makes it easier to determine whether some characteristics of different sites (that offer different programs) influence program impact. However, this approach requires a larger number of participants and, often, cooperation across many governmental or organizational units (St. Pierre and Rossi 2006:667–675).

In a sense, all these choices (black box evaluation or program theory, researcher or stakeholder interests, and so on) hinge on (1) what your real goals are in doing the project and (2) how able you will be in a "research for hire" setting to achieve those goals. Not every agency really wants to know if its programs work, especially if the answer is no. Dealing with such issues and the choices they require is part of what makes evaluation research both scientifically and politically fascinating.

POLICY RESEARCH

Policy research is a process rather than a method: "a process that attempts to support and persuade actors by providing them with well-reasoned, evidence-based, and responsible recommendations for decision making and action" (Majchrzak and Markus 2014:3). Because policy

research often draws on the findings of evaluation research projects and involves working for a client, as is the case in evaluation research, policy researchers confront many of the same challenges as do evaluation researchers. Because policy researchers must summarize and weigh evidence from a wide range of sources, they need to be familiar with each of the methods presented in this book.

The goal of policy research is to inform those who make policy about the possible alternative courses of action in response to some identified problem, their strengths and weaknesses, and their likely positive and negative effects. Reviewing the available evidence may lead the policy researcher to conclude that enough is known about the issues to develop recommendations without further research, but it is more likely that additional research will be needed using primary or secondary sources. Policy researchers Ann Majchrzak and M. Lynne Markus (2014:9) caution those embarking on a policy research project to ensure their plans are

- *Credible:* informed by evidence and unbiased to the extent possible about the pros, cons, and risk of problems and potential interventions
- *Meaningful and engaging* of representatives of stakeholder groups, including policy makers and regulators, and those affected by policy actions, such as customers, suppliers, or service recipients
- *Responsible:* consider a broad spectrum of potential negative consequences of policy change
- *Creative:* recognize needs for new or different solutions
- *Manageable:* doable within the available time and resources

The policy research process begins with the identification of a research question that may focus on the causes of or solutions to the policy problem. For example, Peter Reuter et al. (2010) describe facilitating community adoption of research-based prevention strategies as a policy research question underlying RAND's Drug Policy Research Center (DPRC) Getting to Outcomes project. In this project, prevention scholars, community organizers, and service researchers in the DPRC reviewed the literature and then developed manuals, worksheets, and other resources to help communities adopting prevention programs.

Majchrzak and Markus (2014:22–25) encourage policy researchers to make a clear distinction between the problem they hope to solve and the aspects of the problem they cannot deal with, to review carefully the context in which the problem occurs, to specify clearly why policy change is needed and what possible risks might be incurred by making a change, as well as to develop a causal model of how the policy problem occurs. RAND policy researchers (Reuter et al. 2010:257) found one of the greatest challenges for designing quality policy research is the longer time frame required for producing a good product that involves primary data collection, compared with the short time frame required with most politically driven policy decisions. Stefanie Ettelt et al. (2013) register a similar caution based on their experience in a policy research center funded by the British Department of Health to produce rapid analyses of health care experiences in other countries. Nonetheless, the development of centers like these suggests that policy research is becoming an expected element in policy making. These problems in translating evaluation research results into practice explain why many researchers feel that the best they can hope for is policy "informed" by research evidence rather than "based" on evidence (Fox et al. 2016:13).

ETHICS IN EVALUATION

Evaluation research can make a difference in people's lives while the research is being conducted as well as after the results are reported. Job opportunities, welfare requirements, housing options, treatment for substance abuse, and training programs are each potentially important benefits, and an evaluation research project can change both the type and the availability of such benefits. This direct impact on research participants and, potentially, their families, heightens the attention that evaluation researchers have to give to human subject concerns (Wolf, Turner, and Toms 2009:171). Although the particular criteria that are at issue and the decisions that are judged most ethical vary with the type of evaluation research conducted and the specifics of a particular project, there are always serious ethical as well as political concerns for the evaluation researcher (Boruch 1997:13; Dentler 2002:166).

When program impact is the focus, human subject considerations multiply. What about assigning persons randomly to receive some social program or benefit? One justification for this given by evaluation researchers has to do with the scarcity of these resources. If not everyone in the population who is eligible for a program can receive it because of resource limitations, what could be a fairer way to distribute the program benefits than through a lottery? Random assignment also seems like a reasonable way to allocate potential program benefits when a new program is being tested with only some members of the target recipient population. If the value of a program is not really known, random assignment to an alternative version seems particularly reasonable (Fox et al. 2016:36–37). However, when an ongoing entitlement program is being evaluated and experimental subjects would normally be eligible for program participation, it may not be ethical simply to bar some potential participants from the programs. Instead, evaluation researchers may test alternative treatments or provide some alternative benefit while the treatment is being denied.

There are many other ethical challenges in evaluation research:

- How can confidentiality be preserved when the data are owned by a government agency or are subject to discovery in a legal proceeding?
- Who decides what level of burden an evaluation project may tolerably impose on participants?
- Is it legitimate for research decisions to be shaped by political considerations?
- Must evaluation findings be shared with stakeholders rather than only with policy makers?
- Is the effectiveness of the proposed program improvements really uncertain?
- Will a randomized experiment yield more defensible evidence than the alternatives?
- Will the results actually be used?

The Health Research Extension Act of 1985 (Public Law 99–158) mandated that the Department of Health and Human Services require all research organizations receiving federal funds to have an institutional review board to assess all research for adherence to ethical practice guidelines. We have already reviewed the federally mandated criteria (Boruch 1997:29–33):

- Are risks minimized?
- Are risks reasonable in relation to benefits?
- Is the selection of individuals equitable? (randomization implies this)
- Is informed consent given?
- Are the data monitored?
- Are privacy and confidentiality assured?

Evaluation researchers must consider whether it will be possible to meet each of these criteria long before they even design a study. It is important to inform representatives of all stakeholders about the evaluation before it begins and to consider with them the advantages and disadvantages for potential participants and the type of protections needed (Fox et al. 2016:28–31).

The problem of maintaining subject confidentiality is particularly thorny because researchers, in general, are not legally protected from the requirements that they provide evidence requested in legal proceedings, particularly through the process known as "discovery." However, it is important to be aware that several federal statutes have been passed specifically to protect research data about vulnerable populations from legal disclosure requirements. For example, the Crime Control and Safe Streets Act (28 CFR Part 11) includes the following stipulation:

> Copies of [research] information [about persons receiving services under the act or the subjects of inquiries into criminal behavior] shall be immune from legal process and shall not, without the consent of the persons furnishing such information, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceedings. (Boruch 1997:60)

Ethical concerns must also be given special attention when evaluation research projects involve members of vulnerable populations as subjects. To conduct research on children, parental consent usually is required before the child can be approached directly about the research. Adding this requirement to an evaluation research project can dramatically reduce participation because many parents simply do not bother to respond to mailed consent forms. Sloboda and colleagues' (2009:3) evaluation of the trial TCYL program used an "active consent" procedure for gaining parental consent and student assent: parents and students both had to sign forms before the student could participate; the result was that 58% of the 34,000 eligible seventh-grade students were enrolled in the study. Since nonconsent is likely to be higher among those who are more at risk of substance abuse, the likelihood of identifying program impact can be diminished (Tigges 2003). Other research indicates that use of a "passive consent" procedure—students can participate as long as their parents do not return a form indicating their lack of consent—can result in much higher rates of participation.

Tricia Leakey and her colleagues (2004:511) demonstrated that this problem can be overcome in their evaluation of Project SPLASH (Smoking Prevention Launch Among Students in Hawaii). When the project began in the seventh grade, the researchers gave students project information and a consent card to take home to their parents. A pizza party was then held in every class where at least 90% of the students returned a signed consent card. In subsequent follow-ups in the eighth grade, a reminder letter was sent to parents whose children had previously participated. Classes with high participation rates also received a candy thank you. As you can see in Table 13.5, the result was a very high rate of participation.

TABLE 13.5 ■ Parental Consent Response Rates and Outcomes

Survey	Population Size[a]	Consent to Parents[b]	Consent Returned[c]	Refused Consent[d]	Consent to Participate	tudent Assent "Yes"[e]
7th-grade baseline[f]	4,741	4,728	89.5% (*n* = 4,231)	7.3% (*n* = 310)	92.7% (*n* = 3,921)	99.4% (*n* = 3,716)
8th-grade baseline[g]	4,222	421	58.0% (*n* = 244)	11.9% (*n* = 29)	88.1% (*n* = 215)	99.0% (*n* = 3,235)
8th-grade follow-up[g]	3,703	177	41.8% (*n* = 74)	5.4% (*n* = 4)	94.6% (*n* = 70)	98.7% (*n* = 2,999)

Note: Parents who had refused participation at the previous survey point were again contacted for permission at the next survey point.

[a]Number of students who were enrolled in the program varies over time depending on enrollment and teacher participation rates.

[b]Number of consent forms that were handed out at each time period to new students.

[c]Out of the total number of consent forms distributed.

[d]Out of the total number of consent forms returned.

[e]Out of all students who had parental consent and were present on the day of the survey.

[f]Project staff explained and distributed consent on site.

[g]Teachers explained and distributed consent.

Source: Leakey, Tricia, Kevin B. Lunde, Karin Koga, and Karen Glanz. 2004. "Written Parental Consent and the Use of Incentives in a Youth Smoking Prevention Trial: A Case Study From Project SPLASH." *American Journal of Evaluation* 25:509–523.

When it appears that it will be difficult to meet the ethical standards in an evaluation project, at least from the perspective of some of the relevant stakeholders, modifications should be considered in the study design. Several steps can be taken to lessen any possibly detrimental program impact (Boruch 1997:67–68):

- Alter the group allocation ratios to minimize the number in the untreated control group.
- Use the minimum sample size required to be able to adequately test the results.
- Test just parts of new programs rather than the entire program.
- Compare treatments that vary in intensity (rather than presence or absence).
- Vary treatments between settings rather than between individuals within a setting.

Essentially, each of these approaches limits the program's differential impact during the experiment and so lessens any potential adverse effects on human subjects. It is also important to realize that it is costly to society and potentially harmful to participants to maintain ineffective programs. In the long run, at least, it may be more ethical to conduct an evaluation study in order to improve a program than to let the status quo remain in place (Fox et al. 2016:94).

CONCLUSIONS

Hopes for evaluation research are high: Society could benefit from the development of programs that work well, that accomplish their policy goals, and that serve the people who genuinely need them. At least that is the hope. Unfortunately, there are many obstacles to realizing this hope (Posavac and Carey 1997):

- Because social programs and the people who use them are complex, evaluation research designs can easily miss important outcomes or aspects of the program process.
- Because the many program stakeholders all have an interest in particular results from the evaluation, researchers can be subjected to an unusual level of cross-pressures and demands.
- Because the need to include program stakeholders in research decisions may undermine adherence to scientific standards, research designs can be weakened.
- Because some program administrators want to believe that their programs really work well, researchers may be pressured to avoid null findings or, if they are not responsive, may find their research report ignored. Plenty of well-done evaluation research studies wind up in a recycling bin or hidden away in a file cabinet.
- Because the primary audience for evaluation research reports are program administrators, politicians, or members of the public, evaluation findings may need to be overly simplified, distorting the findings.

The rewards of evaluation research are often worth the risks, however. Evaluation research can provide social scientists with rare opportunities to study a complex social process, with real consequences, and to contribute to the public good. Although they may face unusual constraints on their research designs, most evaluation projects can result in high-quality analysis and publications in reputable social science journals. In many respects, evaluation research is an idea whose time has come. We may never achieve Donald Campbell's (Campbell and Russo 1999) vision of an "experimenting society," in which research is consistently used to evaluate new programs and to suggest constructive changes, but we are close enough to continue trying.

KEY TERMS

Black box evaluation (p. 471)
Cost–benefit analysis (p. 469)
Cost–effectiveness analysis (p. 469)
Efficiency analysis (p. 469)
Evaluability assessment (p. 461)
Feedback (p. 458)
Formative evaluation (p. 465)
Impact evaluation (or analysis)
Inputs (p. 457)
Integrative approach
Needs assessment (p. 459)
Outcomes (p. 458)
Outputs
Policy research (p. 478)
Process evaluation (p. 463)
Program process (p. 457)
Program theory (p. 472)
Social science (researcher) approach
Stakeholder approach
Stakeholders (p. 458)
Summative evaluation (p. 466)
Theory-driven evaluation (p. 472)

HIGHLIGHTS

LO 13.1 Describe the history of evaluation research and its current status.

- Evaluation research is social research that is conducted for a distinct purpose: to investigate social programs.
- It began after the expansion of the federal government during the Great Depression and WWII.
- It became more important with the Great Society programs of 1960s because program evaluation became a frequent requirement when programs were funded.
- The New Jersey Income Maintenance Experiment was the first large-scale experiment to test social policy in action.
- Evaluation research firms declined in the early 1980s as Great Society programs also declined.
- The Government Performance and Results Act of 1993 required some type of evaluation for all government programs.
- The American Evaluation Association was founded in 1986 as a professional organization for evaluation researchers.

LO 13.2 Diagram the evaluation research process as a system.

- The evaluation process can be modeled as a feedback system.
  - Inputs: resources, raw materials, clients, and staff that go into a program
  - Program process: the complete treatment or service delivered by the program
  - Outputs: the services delivered or new products produced by the program
  - Outcomes: the impact of the program on the cases processed
- Feedback: information about service delivery system outputs, outcomes, or operations that is available to any program inputs
- Evaluation research itself is a systematic approach to feedback
- Stakeholders: individuals and groups who have some basis of concern with the program, often setting the research agenda and controlling research findings

LO 13.3 Define each of the types of evaluation research.

- Needs assessment: An attempt to determine whether a new program is needed or an old one is still required.
- Evaluability assessment: A determination about whether a program may be evaluated within the available time and resources.
- Process evaluation (implementation assessment): Evaluation research that investigates the process of service delivery. Can identify the specific aspects of a program that have an impact.
  - Formative evaluation: Process evaluation that is used to shape and refine program operations.
- Impact analysis (also termed summative evaluation): Evaluation research that compares what happened after a program was implemented with what would have happened had there been no program at all. Experimental design is the preferred method for maximizing internal validity.
- Efficiency analysis: Cost–benefit and cost-effectiveness evaluation,
  - Cost–benefit analysis: Are program's financial benefits sufficient to offset the program's costs?
  - Cost-effectiveness analysis: How much does it cost to achieve a given effect?

LO 13.4 Compare the paired alternatives in each of the five design decisions.

- Black box or program theory: Is it important to study how the program gets results?
  - Black box evaluation: If program results are of primary importance, how it works may be of secondary importance.
  - Theory-driven evaluation: When a researcher has sufficient knowledge before the investigation begins, outlining a program theory can help to guide the investigation of a program process in the most productive directions:
    - *Descriptive program theory* specifies impacts that are generated and how this occurs (suggesting a causal mechanism, intervening factors, and context), generally empirically based.
    - *Prescriptive program theory* specifies what ought to be done by the program but has not yet been empirically tested or observed.
- Researcher or stakeholder orientation: Is the primary evaluator of research a set of social scientific peers or a funding agency?
  - Stakeholder approaches encourage researchers to be responsive to stakeholders (a.k.a. responsive evaluation)
    - Utilization-focused evaluation: The evaluator forms a task force of program stakeholders who help to shape the evaluation project so that they are most likely to use its results.
    - Action research (participatory research): Program participants are engaged with the researchers as co-researchers and help to design, conduct, and report the research.
    - Appreciative inquiry: Eliminates the professional researcher altogether in favor of a structured dialogue about needed changes among program participants themselves.
  - Social science approaches: Emphasize the importance of researcher expertise and maintenance of some autonomy to develop the most trustworthy, unbiased program evaluation.
    - Goal free evaluation: Researchers do not know the goals of the program but assess and compare the needs of participants to a wide array of program outcomes to identify unanticipated outcomes.
  - Integrative approaches: Attempt to cover issues of concern to both stakeholders (including participants) and social science evaluators, balancing stakeholder concern with scientific credibility.
- Quantitative or qualitative methods
  - Evaluation research that attempts to identify the effects of a social program typically uses quantitative methods.
  - Qualitative methods are useful for investigating program process, learning how individuals react to treatment, understanding the actual operation of the programs, and understanding more complex social programs.
- Simple or complex outcomes
  - Single outcomes may miss the process of how a program works.
  - Most evaluation research attempts to measure multiple outcomes.
  - Unanticipated outcomes can be missed.
- Groups or individuals

- Individual-based designs: Individuals are randomized or otherwise assigned to conditions being compared in an impact analysis
- Group-based designs: Groups (such as classes or clinics) are randomized or otherwise assigned to conditions being compared in an impact analysis

LO 13.5 Develop an argument for adding the policy research approach to evaluation research projects.

- Policy research is a process rather than a specific method that uses "well-reasoned, evidence-based, responsible recommendations" to inform policy decisions and action steps.
- Focus may be on causes or solutions to a social problem.
- Need to carefully review context in which problem occurs.
  - Engage in this to ensure that plans for research are credible, meaningful and engaging, responsible, creative, and manageable.

LO 13.6 Defend the ethics of research evaluations of social programs.

- Evaluation research raises complex ethical issues.
  - The direct impact on participants and their families through social programs heightens the attention to human subjects concerns.
  - Needs assessments, evaluability assessments, process analysis, and cost–benefit analysis have few special ethical considerations.
  - When program impact is the focus, human subjects problems multiply:
    - When different participants receive different treatments, equitable distribution of benefits may become a concern
    - Preserving confidentiality may be difficult
    - The potential burden imposed by participating in research should be considered
    - Research designs are often shaped by politics
  - Federally mandated IRBs must assess all research for adherence to ethical practice guidelines:
    - Minimize risks to participants
    - Allow only risks that are reasonable in relation to benefits
    - Select individuals equitably (randomization implies this)
    - Ensure informed consent
  - To lessen any detrimental program impact
    - Minimize the number in the control group (but consult a statistician first)
    - Use a minimum sample size that still allows the needed statistical tests
    - Test only new parts of the program, not the entire program
    - Compare treatments that vary in intensity rather than presence and absence
    - Vary treatments between settings rather than among individuals in a single setting

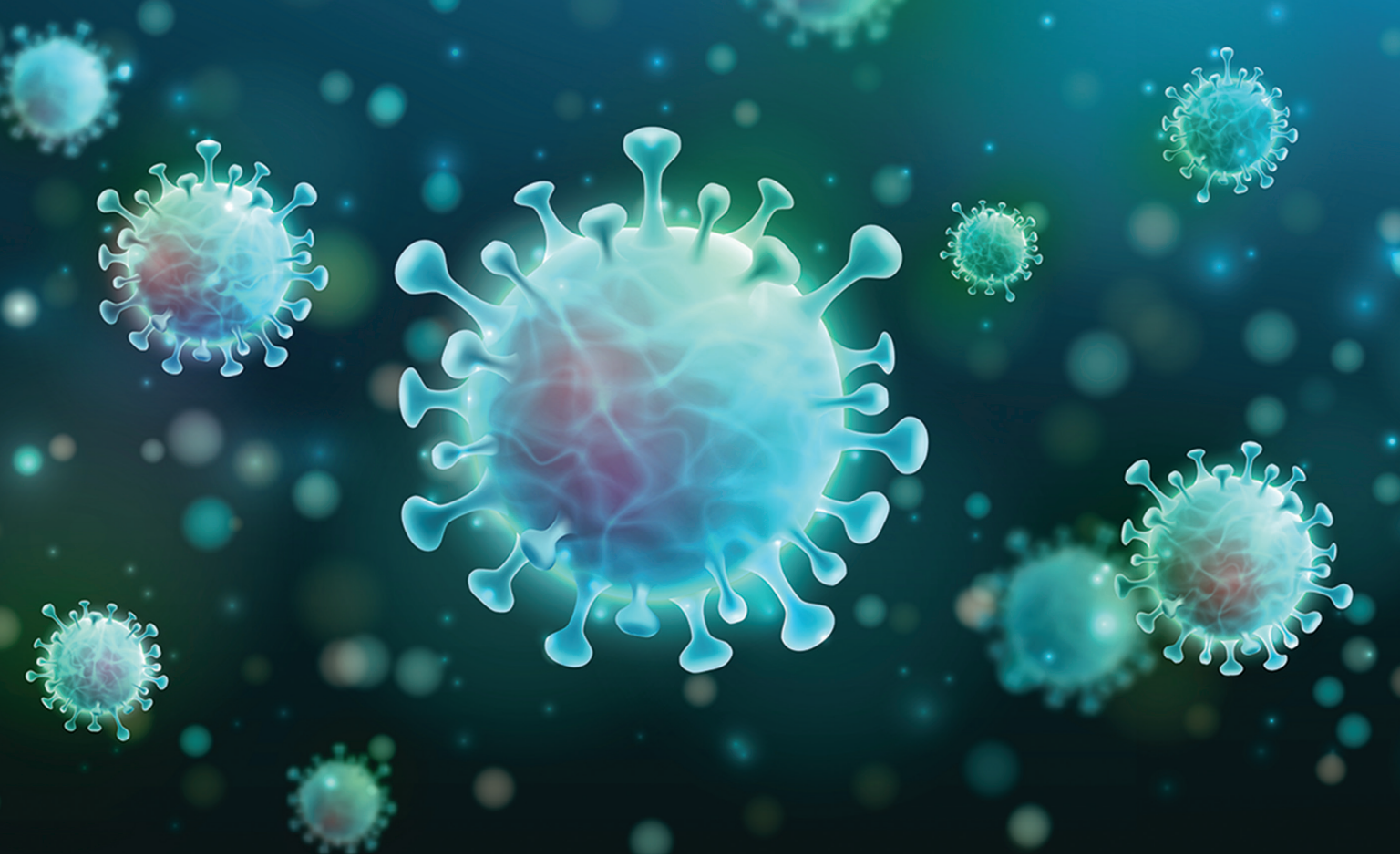
DISCUSSION QUESTIONS

1. Would you prefer that evaluation researchers use a stakeholder or a social science approach? Compare and contrast these perspectives, and list at least four arguments for the one you favor.

2. Propose a randomized experimental evaluation of a social, medical, or educational program with which you are familiar. Possibilities could range from a job training program to a community health initiative or even a college program. Include in your proposal a description of the program and its intended outcomes. Discuss the strengths and weaknesses of your proposed design.
3. How would you describe the contents of the "black box" of program operations? What "program theory" would specify how the program (in Question 2) operates?
4. What would be the advantages and disadvantages of using qualitative methods to evaluate this program? What would be the advantages and disadvantages of using quantitative methods? Which approach would you prefer and why?

SPSS EXERCISES

1. Neighborhood and school integration has often been a focus of government social policy. Does the racial composition of a neighborhood have any association with attitudes related to racial issues? Although we cannot examine the effects of social policies or programs directly in the General Social Survey (GSS) data, we can consider the association between neighborhood racial composition and attitudes related to race. The variable RACLIVE_2 indicates whether the respondent lives in a racially integrated neighborhood. Request its frequency distribution as well as those for several attitudes related to race: AFFRMACT_2, WRKWAYUP_2, and HELPBLK_2.
2. Do attitudes vary with the experience of living in a racially integrated neighborhood? Request the cross-tabulation of the variables used in Step 1, AFFRMACT_2, WRKWAYUP_2, HELPBLK_2 by RACLIVE_2 by RACE2020 (request percentages on the column totals). Read the tables and explain what they tell us about attitudes and neighborhoods and race. Does the apparent effect of racial integration for whites vary with the different attitudes? How would you explain this variation in these "multiple outcomes"?
3. What other attitudes differ between whites who live in integrated and segregated neighborhoods? Review the GSS2020 variable list to identify some possibilities and request cross-tabulations for these variables and race Do you think these differences are more likely to be a consequence of a racially integrated neighborhood experience or a cause of the type of neighborhood that people choose to live in? Explain.



©iStockphoto.com/fotomay

14 USING SECONDARY DATA AND BIG DATA

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

14.1 Explain the steps involved in using ICPSR

14.2 Describe at least four other major sources of secondary data for social scientists

14.3 Defend the use of secondary data, considering the challenges it poses

14.4 Assess the strengths and weaknesses of research using Big Data

14.5 Plan an IRB presentation about the ethics of a Big Data research project

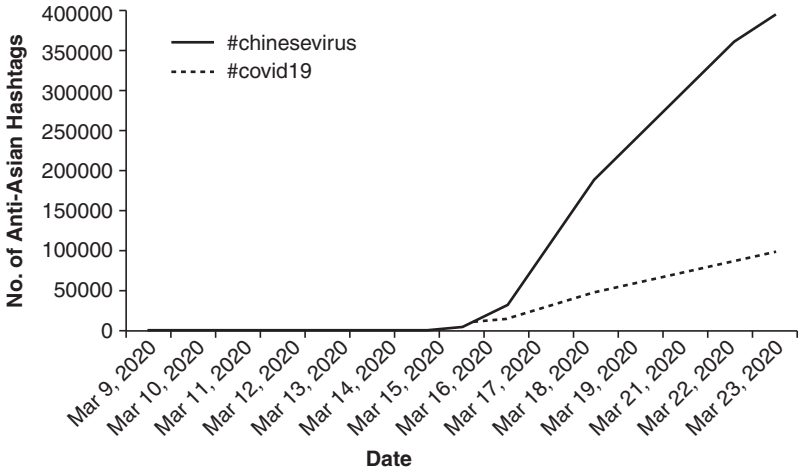
> **D**isease names really do matter. . . . We've seen certain disease names provoke a backlash against members of particular religious or ethnic communities. (WHO 2015)

This statement by the World Health Organization (WHO 2015) drew on recent experience with new disease names like "Middle East Respiratory Syndrome," "Rift Valley fever," and "swine flu" being followed by a backlash against a particular religious or ethnic community or needless slaughter of animals. The WHO recommended naming diseases based on their symptoms or the underlying pathogen (such as salmonella). Just five years later, the emergence of SARS-Cov-2 (severe acute respiratory syndrome coronavirus 2) in Wuhan, China, put the WHO's idea to the test.

Yulin Hswen in the Department of Epidemiology and Biostatistics, University of California San Francisco, and her collaborators Xiang Xu, Anna Hing, Jared B. Hawkins, John S. Brownstein, and Gilbert C. Gee (2021) decided to investigate the association between the name used for the pandemic disease and anti-Asian sentiment during the first weeks of the pandemic's spread in the United States. President Trump's first tweet using the phrase "Chinese Virus," on March 16, 2020, increased attention to this research question. Hashtags summarize the intention of a tweet by indicating the audience that the user seeks to reach, so Dr. Hswen's team focused on the 668,597 tweets dated from one week before to one week after the president's "Chinese Virus" (May 9 to 23, 2020) tweet that included the hashtag "#chinesevirus" or "#covid19." As you can see in Figure 14.1, the number of anti-Asian hashtags also used in these tweets rose dramatically after March 16th, with most of the rise being in tweets that included the #chinesevirus hashtag (Hswen et al. 2021).

This chapter's focus is on methods of research that allow investigations of the social world with data available electronically. The data collected in many large funded social research projects using surveys and other basic social research designs are available in a digitized format on the web, where they can be downloaded or, in some cases, analyzed directly. I will first identify many of the sources for such "secondary data" and explain how to obtain data from these sources. I will also review some easy-to-overlook problems that create challenges for those using secondary data. Social researchers can also obtain Big Data that is collected electronically but represents frequent physical activities in the social world, like calls to 911 or sales prices of homes. I will describe the use of such Big Data in research and also introduce "digital methods" that involve analysis of social interaction that occurs online. This is the approach used by researchers like Dr. Hswen, who study social behavior represented by tweets or other forms of electronic communication.

FIGURE 14.1 ■ Timeline of Anti-Asian Twitter Hashtags Under #covid19 and #chinesevirus, March 9-23, 2020



Source: Hswen, Yulin, Xiang Xu, Anna Hing, Jared B. Hawkins, John S. Brownstein, and Gilbert C. Ghee. 2021. "Association of '#covid19' versus '#chinesevirus' with Anti-Asian Sentiments on Twitter: March 9-23, 2020." *American Journal of Public Health* 111(5):961.

SECONDARY DATA ANALYSIS

Secondary data analysis is the method of using preexisting data in a different way or to answer a different research question than intended by those who collected the data. The most common sources of **secondary data**—previously collected data that are used in a new analysis—are social science surveys and data collected by government agencies, often with survey research methods. It is also possible to reanalyze data that have been collected in experimental studies or with qualitative methods. Even a researcher's reanalysis of data they collected previously qualifies as secondary analysis if it is employed for a new purpose or in response to a methodological critique.

Overview

Thanks to the data collected by social researchers, governments, and organizations over many years, secondary data analysis has become the research method used by many contemporary social scientists to investigate important research questions. Why consider secondary data? (1) Data collected in previous investigations are available for use by other social researchers on a wide range of topics. (2) Available data sets often include many more measures and cases and reflect more rigorous research procedures than another researcher will have the time or resources to work with in a new investigation. (3) Much of the groundwork involved in creating and testing measures with the data set has already been done. (4) Most important, most funded social science research projects collect data that can be used to investigate new research questions that were not considered by the primary researchers who collected the data. Analyzing secondary data, then, is nothing like buying "used goods"!

Secondary data analysis has been an important social science methodology since the earliest days of social research, whether when Karl Marx (1967) reviewed government statistics in the Reading Room of the British Library during the 1850s to 1870s or when Émile Durkheim (1966) analyzed official government cause-of-death data for his study of suicide rates throughout

Europe in the late 19th century. With the advent of modern computers and, even more important, the Internet, secondary data analysis has become an increasingly accessible social research method. Literally, thousands of large-scale data sets are now available for the secondary data analyst, often with no more effort than the few commands required to download the data set. A number of important data sets can even be analyzed directly on the web by users who lack their own statistical software.

There are many sources of data for secondary analysis within the United States and internationally. These sources range from data compiled by governmental units and private organizations for administrative purposes, which are subsequently made available for research purposes, to data collected by social researchers seeking to answer one research question that are then made available for reanalysis to answer different research questions. Many important data sets are collected for the specific purpose of facilitating secondary data analysis. Government units, from the Census Bureau to the U.S. Department of Housing and Urban Development; international organizations, such as the United Nations, the Organization for Economic Co-operation and Development (OECD), and the World Bank; and internationally involved organizations, such as the Central Intelligence Agency (CIA), sponsor a substantial amount of social research that is intended for use by a broader community of social scientists. The National Opinion Research Center (NORC), with its General Social Survey (GSS) is a good example of academically based research that is intended to gather data for social scientists to use in analyzing a range of social science research questions.

Many social scientists who have received funding to study one research question have subsequently made the data they collect available to the broader social science community for investigations of other research questions. Many of these data sets are available from a website maintained by the original research organization, often with some access restrictions. Examples include the Add Health study conducted at the University of North Carolina Population Center, the University of Michigan's Health and Retirement Study, and the United Nations University's World Income Inequality Database.

What makes secondary data analysis such an exciting and growing option today are the considerable resources being devoted to expanding the amount of secondary data and to making it available to social scientists. For example, the National Data Program for the Social Sciences, funded in part by the National Science Foundation, sponsors the ongoing General Social Survey to make current data on a wide range of research questions available to social scientists (https://gss.norc.org/). Since 1985, the GSS has participated in an International Social Survey Program that generates comparable data from 47 countries around the world (www.issp.org). Another key initiative is the Data Preservation Alliance for the Social Sciences (Data-PASS), initiated by the Library of Congress in 2004 as a part of the National Digital Preservation Program (http://data-pass.org/). This project is designed to ensure the preservation of digitized social science data. As represented in efforts by the **Inter-university Consortium for Political and Social Research (ICPSR)** at the University of Michigan (http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/preservation/policies/index.html), this digital preservation policy combines the efforts of other major social research organizations, including the Roper Center for Public Opinion Research at the University of Connecticut; the Howard W. Odum Institute for Research in Social Sciences at the University of North Carolina, Chapel Hill; the Henry A. Murray Research Archive and the Harvard-MIT Data Center at Harvard University; and the Electronic and Special Media Records Service Division of the U.S. National Archives and Records Administration.

Fortunately, you do not have to google your way around the web to find all these sources on your own. Many websites provide extensive collections of secondary data. Chief among these is

the ICPSR website at the University of Michigan (https://www.icpsr.umich.edu/web/pages/). The University of California at Berkeley's Survey Documentation and Analysis (SDA) archive provides several data sets from national omnibus surveys, as well as from U.S. Census microdata, from surveys on racial attitudes and prejudice, and from several labor and health surveys (https://sda.berkeley.edu/). The National Archive of Criminal Justice Data is an excellent source of data in criminal justice (https://nicic.gov/national-archive-criminal-justice-data-nacjd), although, like many other data collections—including key data from the U.S. Census—it is also available through ICPSR. Much of the statistical data collected by U.S. federal government agencies can be accessed through the consolidated USA.gov website, https://www.usa.gov/statistics.

In this section, I will introduce the method of secondary data analysis by describing the ICPSR, the most comprehensive source of data useful in social science research.

Inter-university Consortium for Political and Social Research

The University of Michigan's ICPSR is the premier source of secondary data useful to social science researchers (https://www.icpsr.umich.edu/web/pages/index.html). ICPSR was founded in 1962 and now includes more than 750 colleges and universities and other research organizations throughout the world. ICPSR archives the most extensive collection of social science data sets in the United States outside the federal government: More than 16,899 studies are represented in from around the world and from sources ranging from U.S. government agencies, such as the Census Bureau, to international organizations, such as the United Nations, social research organizations, such as the National Opinion Research Center, and individual social scientists who have completed funded research projects. More than 100,000 publications using data available at ICPSR are included in the searchable ICPSR bibliography.

Types of Data Available From ICPSR

Survey data sets obtained in the United States and in many other countries that are stored at ICPSR provide data on topics ranging from elite attitudes to consumer expectations. For example, data collected in the British Social Attitudes Survey in 1998, designed by the University of Chicago's National Opinion Research Center, are available through ICPSR (go to the ICPSR website, www.icpsr.umich.edu, choose "Find Data," and search for "ICPSR 3101"). Data collected in a monthly survey of Spaniards' attitudes, by the Center for Research on Social Reality (Spain) Survey, are also available (see "ICPSR 6964"). Survey data from Russia, Germany, and other countries can also be found in the ICPSR collection.

Do you have an interest in events and interactions between nations, such as threats of military force? A data set collected by Charles McClelland includes characteristics of 91,240 such events (study no. 5211). The history of military interventions in nations around the world between 1946 and 1988 is coded in a data set developed by Frederic Pearson and Robert Baumann (study no. 6035). This data set identifies the intervener and target countries, the starting and ending dates of military intervention, and a range of potential motives (such as foreign policies, related domestic disputes, and the pursuit of rebels across borders).

Census data from other nations are also available through ICPSR, as well as directly through the Internet. In the ICPSR archives, you can find a data set from the Statistical Office of the United Nations on the 1966 to 1974 population of 220 nations throughout the world (study no. 7623). More current international population data are available through data sets from a variety of sources, such as the study of indicators of globalization from 1975 to 1995 (study no. 4172). (See also the later description of the Eurobarometer Survey Series.) More than 3,000 data

sets from countries outside the United States are available through ICPSR's International Data Resource Center.

Obtaining Data From ICPSR

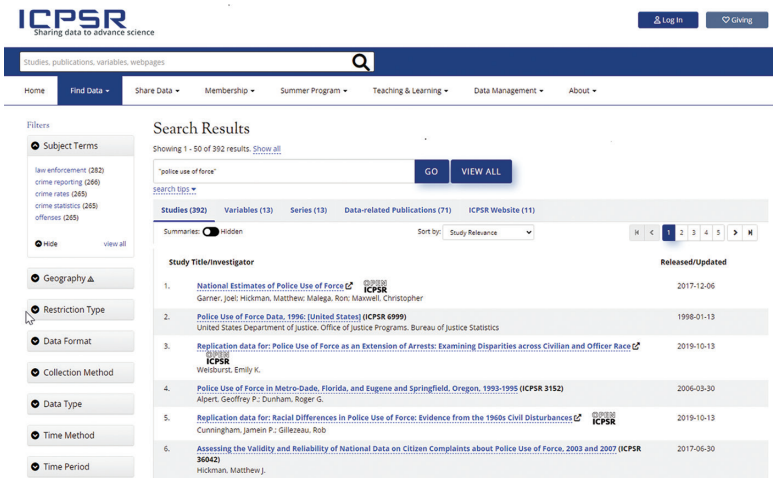
The data sets archived by ICPSR are available for downloading directly from the ICPSR website, www.icpsr.umich.edu. ICPSR makes data sets obtained from government sources available directly to the public, although many other data sets are available only to individuals at the colleges and universities around the world that have paid the fees required to join ICPSR. The availability of some data sets is restricted because of confidentiality issues (see the section later in this chapter on research ethics); to use them, researchers must sign a contract and agree to certain conditions (see http://www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted/index.html).

You begin a search for data in the ICPSR archives at www.icpsr.umich.edu. You can search the data archives for a specific topic, for specific studies (identified by study number [it will be the first listed in the search results] or title), as well as for studies by specific investigators (this would be a quick way to find the data set contributed by Lawrence W. Sherman and Richard A. Berk from their research, discussed in Chapter 2, on the police response to domestic violence).

Figure 14.2 displays the results of a search for data sets on "police use of force": a list of 392 data sets from studies that involved research on the police use of force and that are available through ICPSR. For most data sets, you can obtain a description, the files that are available for downloading, and a list of "related literature"—that is, reports and articles that use the listed data set.

As you review the list of datasets, you should carefully read the study descriptions, scan at least some of the publications or reports resulting from the studies, and examine the lists of variables. Consider how well each study matches your research interest or specific research question. If there is more than one top contender, consider downloading more than one dataset and exploring the data in more depth.

FIGURE 14.2 ■ Search Screen: Police Use of Force Search Results



Source: Inter-university Consortium for Political and Social Research.

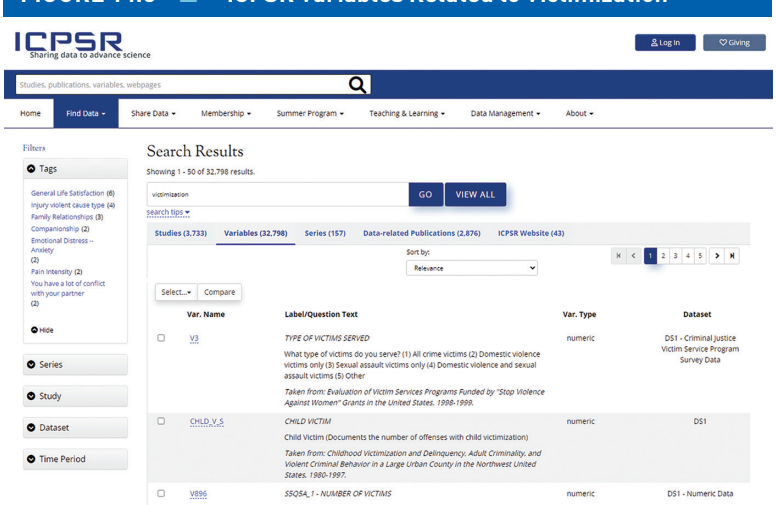
When you click on the "Download" option, you are first asked to enter your e-mail address and password (ICPSR also has the option that allows sign-in with a Facebook account). What you enter will determine which data sets you can access; if you are not at an ICPSR member institution, you will be able to download only a limited portion of the data sets—mostly those from government sources. If you are a student at a member institution, you will be able to download most of the data sets directly, although you may have to be using a computer that is physically on your campus to do so.

If you prepare your own paper based on an analysis of ICPSR data, be sure to include a proper citation. Here's an example from ICPSR itself (https://www.icpsr.umich.edu/web/pages/datamanagement/citations.html):

Schneider, Barbara, and Waite, Linda J. The 500 Family Study [1998-2000: United States]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2008-06-03. https://doi.org/10.3886/ICPSR04549.v1

You can also search the ICPSR database for specific variables from among the more than 5 million included in about three quarters of its collection and then identify the various studies in which these variables have appeared. Figure 14.3 displays one segment of the results of searching for variables related to "victimization." A total of 32,772 names of variables in a multitude of studies were obtained. Reviewing some of these results can suggest additional search strategies and alternative databases to consider.

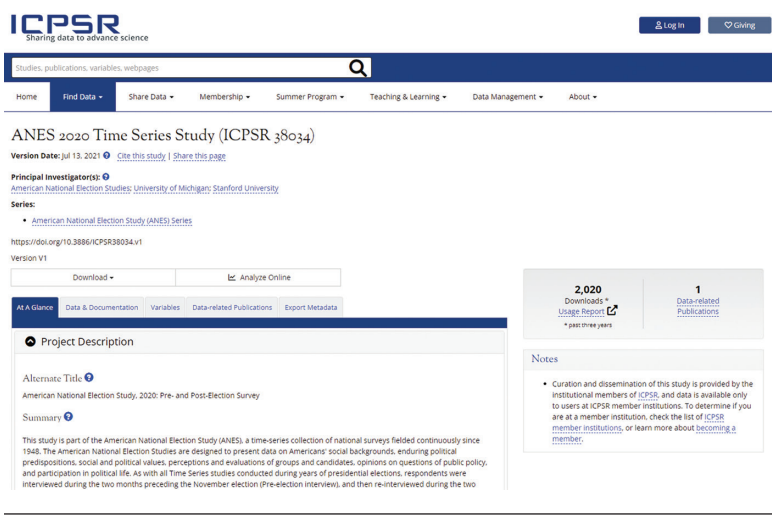
FIGURE 14.3 ■ ICPSR Variables Related to Victimization



Source: Inter-university Consortium for Political and Social Research.

Some of the data sets are also offered with the option to "Analyze Online." If you have this option, you can immediately inspect the distributions of responses to each question in a survey and examine the relationship between variables, without having any special statistical program of your own. To use this option, you must log in to ICPSR through your university library site or generate your own username and password. After doing so, you can choose the "Analyze Online" option for a data set that has this option. For example, after identifying my desired Data Format as "Online analysis" on the left of the ICPSR search screen, I searched for studies indexed with the words "coronavirus pandemic." In Figure 14.4, you'll find the screen displaying the information provided about one of the studies that matched my request, the "ANES 2020 Time Series Study."

FIGURE 14.4 ■ ICPSR Online Analysis: Codebook Information and Statistical Options


ICPSR
Sharing data to advance science
Log In
Giving
Studies, publications, variables, webpages
Home
Find Data
Share Data
Membership
Summer Program
Teaching & Learning
Data Management
About
ANES 2020 Time Series Study (ICPSR 38034)
Version Date: Jul 13, 2021
Cite this study | Share this page
Principal Investigator(s):
American National Election Studies; University of Michigan; Stanford University
Series:
American National Election Study (ANES) Series
https://doi.org/10.3886/ICPSR38034.v1
Version V1
Download
Analyze Online
At A Glance
Data & Documentation
Variables
Data-related Publications
Export Metadata
Project Description
Alternate Title
American National Election Study, 2020: Pre- and Post-Election Survey
Summary
This study is part of the American National Election Study (ANES), a time-series collection of national surveys fielded continuously since 1948. The American National Election Studies are designed to present data on Americans' social backgrounds, enduring political predispositions, social and political values, perceptions and evaluations of groups and candidates, opinions on questions of public policy, and participation in political life. As with all Time Series studies conducted during years of presidential elections, respondents were interviewed during the two months preceding the November election (Pre-election interview), and then re-interviewed during the two
2,020
Downloads *
Usage Report
* past three years
1
Data-related Publications
Notes
Curation and dissemination of this study is provided by the institutional members of ICPSR, and data is available only to users at ICPSR member institutions. To determine if you are at a member institution, check the list of ICPSR member institutions, or learn more about becoming a member.

Source: Inter-university Consortium for Political and Social Research.

If you click on the option to conduct an online analysis, you'll need to log in as a registered ICPSR user. If your school subscribes to ICPSR, you can register in a separate step. You can then find variables in the study and request frequencies of any variable or crosstabs of one variable by another. Have fun! This approach to analysis with secondary data can jump-start your work (and may jump-start some interesting discussion as well). An online analysis option is also starting to appear at other websites that offer secondary data.

ICPSR also catalogs reports and publications containing analyses that have used ICPSR data sets since 1962—more than 101,000 citations were in this archive on April 18, 2022. This superb resource provides an excellent starting point for the literature search that should precede a secondary data analysis. In most cases, you can learn from detailed study reports a great deal about the study methodology, including the rate of response in a sample survey and the reliability of any indexes constructed. Published articles provide examples of how others have described the study methodology as well as research questions that have already been studied with the data set and issues that remain to be resolved. You can search this literature at the ICPSR site simply by entering the same search terms you used to find data sets or by entering the specific study number of the data set on which you have focused. Don't start a secondary analysis without searching "Data-Related Publications" and reviewing at least some of the reports and publications.

Of course, even if you are using ICPSR, you shouldn't stop your review of the literature with the sources listed on the ICPSR site. Conduct a search in SocINDEX or another bibliographic database to learn about related studies that used different databases (see Chapter 2).

SECONDARY DATA FROM GOVERNMENT AND OTHER SOURCES

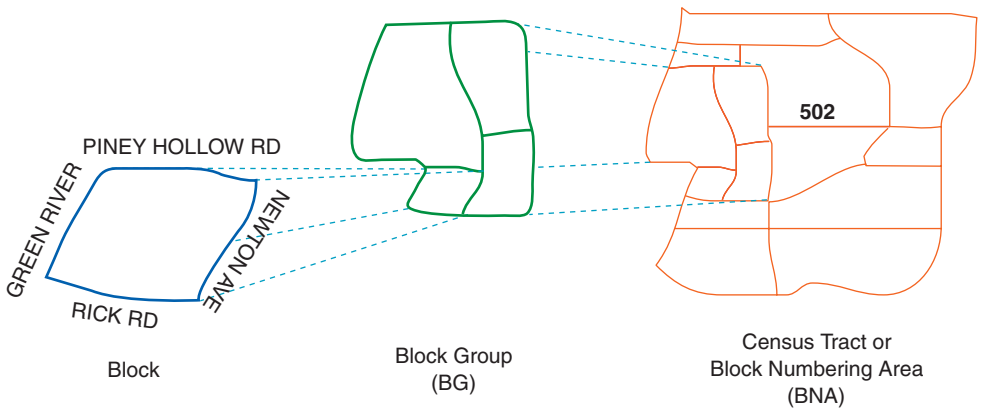
In this section, I will describe other major sources of online data. Many are available from the U.S. Census and other government agencies, including the U.S. Census Bureau's Current Population Survey (CPS) and its Survey of Manufactures and the Bureau of Labor Statistics' Consumer Expenditure Survey. These government data sets typically are quantitative; in fact, the term *statistics*—state-istics—is derived from this type of data.

U.S. Census Bureau

The U.S. government has conducted a census of the population every 10 years since 1790; since 1940, this census has also included a census of housing (see Chapter 5). This decennial Census of Population and Housing is a rich source of social science data (Lavin 1994). The Census Bureau's monthly CPS provides basic data on labor force activity that is then used in U.S. Bureau of Labor Statistics reports. The Census Bureau also collects data on agriculture, manufacturers, construction and other business, foreign countries, and foreign trade.

The U.S. Census of Population and Housing aims to survey one adult in every household in the United States. The basic *complete-count* census contains questions about household composition as well as ethnicity and income. More questions are asked in a longer form of the census that is administered to a sample of the households. The separate census of housing characteristics is conducted at the same time (Rives and Serow 1988:15). Participation in the census is required by law, and confidentiality of the information obtained is mandated by law for 72 years after collection. Census data are reported for geographic units, including states, metropolitan areas, counties, census tracts (small, relatively permanent areas within counties), and even blocks (see Figure 14.5). These different units allow units of analysis to be tailored to research questions. Census data are used to apportion seats in the U.S. House of Representatives and to determine federal and state legislative district boundaries, as well as to inform other decisions by government agencies.

FIGURE 14.5 ■ Census Small-Area Geography



Source: U.S. Bureau of the Census (1994:8).

The U.S. Census website (www.census.gov) provides much information about the nearly 100 surveys and censuses that the Census Bureau directs each year, including direct access to many statistics for particular geographic units. Census data can be searched directly in the interactive data retrieval system at https://data.census.gov/cedsci/. The catalog of ICPSR (www.icpsr.umich.edu/icpsrweb/ICPSR) also lists many census reports. You can review U.S. Census Tools and Apps for data retrieval at https://www.census.gov/data/data-tools.html. An even more accessible way to use U.S. Census data is through the website maintained by the Social Science Data Analysis Network, at www.ssdan.net. Check out the DataCounts! options.

Integrated Public Use Microdata Series

Individual-level samples from U.S. Census data, as well as census files from 102 other countries are available through the Integrated Public Use Microdata Series (IPUMS) at the University

of Minnesota's Minnesota Population Center (MPC) (https://www.ipums.org). These data are prepared in an easy-to-use format that provides consistent codes and names for all the different samples. You must register to download data, but the registration is free.

This exceptional resource offers samples of the U.S. population selected from federal censuses, as well as results of the Census Bureau's annual American Community Survey from 2000 to the present. Each sample is independently selected so that individuals are not linked between samples. In addition to basic demographic measures, variables in the U.S. samples include educational, occupational, and work indicators; respondent income; disability status; immigration status; veteran status; and various household characteristics, including family composition and dwelling characteristics. Survey data are also organized in the categories of health, higher education, time use, demography and health, and the environment. The international samples include detailed characteristics from more than one billion person records in countries ranging from France and Mexico to Kenya and Vietnam. Data are also available from global health research, environmental data, time-use studies, on higher education, and on specific geographic units.

Bureau of Labor Statistics

Another good source of data is the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor, which collects and analyzes data on employment, earnings, prices, living conditions, industrial relations, productivity and technology, and occupational safety and health (U.S. Bureau of Labor Statistics 1991; 1997b). Some of these data are collected by the U.S. Census Bureau in the monthly Current Population Survey (CPS); other data are collected through surveys of establishments (U.S. Bureau of Labor Statistics 1997a).

The CPSprovides a monthly employment and unemployment record for the United States, classified by age, sex, race, and other characteristics. The CPS uses a stratified random sample of about 60,000 households (with separate forms for about 120,000 individuals) (https://www.bls.gov/cps/). Detailed questions are included to determine the precise labor force status (whether they are currently working or not) of each household member over the age of 16. Statistical reports are published each month in the BLS's *Monthly Labor Review* and can also be inspected at its website (https://www.bls.gov/mlr/).

BLS also sponsors the National Longitudinal Surveys (NLS), which have been surveying children, youth, and adults at various ages since the mid-1960s. Measures range from education and military experience to income, mobility, health, nutrition, physical activity, fertility and sexual activity, parenting, attitudes and behaviors (https://www.bls.gov/nls/).

Other Government Sources

Many more datasets useful for historical and comparative research have been collected by federal agencies and other national government organizations. The National Technical Information Service (NTIS) of the U.S. Department of Commerce maintains a reports library that collects many reports based on these data resources.

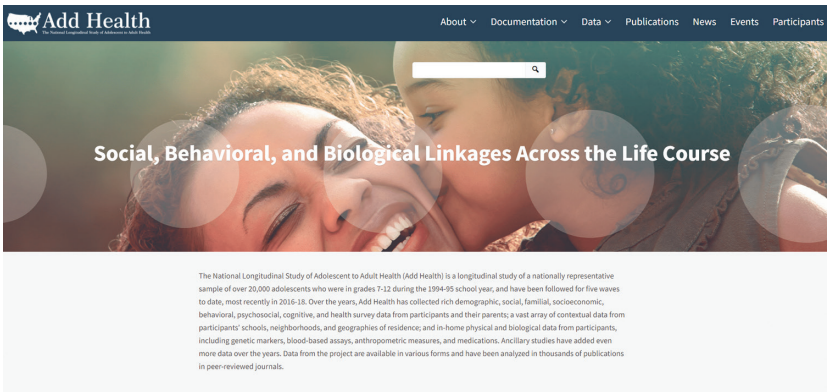
State governments are also increasingly providing datasets that researchers can download. For instance, New York's data resources may be found at https://data.ny.gov, and the state of Missouri makes census data available at http://mcdc.missouri.edu.

Other Data Sources

Many researchers who have received funding to investigate a wide range of research topics make their data available on websites where they can be downloaded by other researchers for

secondary data analyses. One of the largest, introduced earlier, is the Add Health study, funded at the University of North Carolina by the National Institute of Child Health and Human Development (NICHD) and 23 other agencies and foundations to investigate influences on adolescents' health and risk behaviors (https://addhealth.cpc.unc.edu/) (see Figure 14.6). The study began in 1994–1995 with a representative sample of more than 90,000 adolescents who completed questionnaires in school and more than 20,000 who were interviewed at home. This first wave of data collection has been followed by four more, resulting in longitudinal data currently available for more than 20 years. Related data are available from a survey of parents in 2015–2017 and from genetic data collected in 2001–2002 and 2008 (from saliva samples).

FIGURE 14.6 ■ Add Health Website


Add Health
About
Documentation
Data
Publications
News
Events
Participants
Social, Behavioral, and Biological Linkages Across the Life Course
The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a longitudinal study of a nationally representative sample of over 20,000 adolescents who were in grades 7-12 during the 1994-95 school year, and have been followed for five waves to date, most recently in 2016-18. Over the years, Add Health has collected rich demographic, social, familial, socioeconomic, behavioral, psychosocial, cognitive, and health survey data from participants and their parents; a vast array of contextual data from participants' schools, neighborhoods, and geographies of residence; and in-home physical and biological data from participants, including genetic markers, blood-based assays, anthropometric measures, and medications. Ancillary studies have added even more data over the years. Data from the project are available in various forms and have been analyzed in thousands of publications in peer-reviewed journals.

Source: Add Health (https://addhealth.cpc.unc.edu/).

Another significant data source, the Health and Retirement Study (HRS), also including genetic data, began in 1992 with funding from the National Institute on Aging (NIA) (http://hrsonline.isr.umich.edu). The University of Michigan oversees HRS interviews every two years with more than 22,000 Americans over the age of 50. To investigate family experience change, researchers at the University of Wisconsin designed the National Survey of Families and Households (http://www.ssc.wisc.edu/nsfh/). With funding from both NICHD and NIA, researchers interviewed members of more than 10,000 households in three waves, from 1987 to 2003. The Fragile Families & Child Wellbeing Study at Princeton offers a multidimensional collection of data from medical records, surveys of caregivers, parents, and children, and observation as it follows 5,000 children born in large U.S cities—many of them to unmarried parents—between 1998 and 2000 (https://fragilefamilies.princeton.edu/).

The Roper Center archives public opinion poll data, with more than 23,000 U.S. and international data sets and almost 700,000 questions (and answers) from U.S. polls of adults (https://ropercenter.cornell.edu). Another noteworthy example, among many, is the Detroit Area Studies, with annual surveys between 1951 and 2004 on a wide range of personal, political, and social issues (http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/151).

Societal Experts Action Network (SEAN)

Social scientists mobilized quickly after the outbreak of COVID-19 to answer key questions about the impact of the pandemic and to develop evidence-based recommendations for policy makers. In the United States, the National Academies of Sciences, Engineering, and

Medicine received funding from the National Science Foundation and the Alfred P. Sloan Foundation to establish the Societal Experts Action Network (SEAN) as a resource to support responses at the local, state, and national levels (https://www.nationalacademies.org/our-work/societal-experts-action-network).

The SEAN Survey Archive provides summaries of survey results throughout the pandemic and also makes available many of the datasets for further analysis: https://covid-19.parc.us.com/client/index.html#/ (enter Search/Keyword: data and check "Other").

Harvard's Dataverse

Harvard University's Henry A. Murray Research Archive (https://murray.harvard.edu) has developed a remarkable collection of social science research data sets that are now made available through Harvard's collaborative secondary data project at its Institute for Quantitative Social Science (IQSS) (https://dataverse.harvard.edu). As of December 2021, Harvard's Dataverse Project cataloged more than 140,300 datasets (54,843 in the social sciences), including 1,411 collected about some aspect of the COVID-19 pandemic. You can search data sets in the Dataverse collection by title, abstract, keywords, and other fields; if you identify a data set that you would like to analyze, you must then apply to be given access.

International Data Sources

Comparative researchers (using techniques we will discuss in Chapter 15) and those conducting research in specific other countries can find data sets on the population characteristics, economic and political features, and political events of many nations. Some of these data sets are available from U.S. government agencies. For example, the Social Security Administration reports on the characteristics of social security throughout the world (Wheeler 1995). This comprehensive source classifies nations by their type of social security program and provides detailed summaries of the characteristics of each nation's programs. Current information is available online at https://www.ssa.gov/policy/docs/progdesc/ssptw/. More recent data are organized by region. A broader range of data is available in the *World Handbook of Political and Social Indicators*, with political events and political, economic, and social data coded from 1948 to 1982 (www.icpsr.umich.edu, study no. 7761) (Taylor and Jodice 1986).

The European Commission administers the Eurobarometer Survey Series at least twice yearly across all the member states of the European Union. The survey monitors social and political attitudes and reports are published regularly online at https://www.gesis.org/eurobarometer-data-service/survey-series/. Case-level Eurobarometer survey data are stored at ICPSR. The United Nations University makes available a World Income Inequality Database from ongoing research on income inequality in developed, developing, and transition countries (https://www.wider.unu.edu/data).

The Council of European Social Science Data Archives (CESSDA, CESSDA Data Catalogue) catalogs data from studies in a wide range of European nations (Dale, Wathan, and Higgins 2008:521). In April 2022, the CESSDA provided more than 31,000 studies in English out of a total of almost 41,000 in any language. Some are available for downloading.

Qualitative Data Sources

Far fewer qualitative data sets are available for secondary analysis, but the number is growing. European countries, particularly England, have been at the forefront of efforts to promote archiving of qualitative data. The United Kingdom's Economic and Social Research Council

established the Qualitative Data Archiving Resource Center at the University of Essex in 1994 (Heaton 2008:507). Now part of the Economic and Social Data Service, the UK Data Service QualiBank (https://discover.ukdataservice.ac.uk//QualiBank) provides access to 1,439 files from qualitative datasets (as of April 18, 2022). After registering at the UK Data Service site, users can browse or search directly interview transcripts and other materials from many qualitative and mixed-methods studies. Some studies are available for direct downloading, but others are restricted to users in the United Kingdom or according to other criteria.

In the United States, the ICPSR collection includes an expanding number of studies containing at least some qualitative data or measures coded from qualitative data. Studies range from transcriptions of original handwritten and published materials relating to infant care and childcare from the beginning of the 20th century to World War II (LaRossa 1995) to transcripts of open-ended interviews with high school students involved in violent incidents (Lockwood 1996). Harvard University's Dataverse (https://dataverse.harvard.edu) also includes qualitative studies and mixed-methods studies that contain at least some qualitative data.

The most unique source of qualitative data available for researchers in the United States is the Human Relations Area Files (HRAF) at Yale University, described in Chapter 15. The HRAF (https://hraf.yale.edu/) has made anthropological reports available for international cross-cultural research since 1949 and now contains in its online database (eHRAF) more than 680,000 pages of information on more than 310 different cultural, ethnic, religious, and national groups (https://hraf.yale.edu/faq/). If you are interested in cross-cultural research, it is well worth checking out the HRAF and exploring access options (reports can be accessed and searched online by those at affiliated institutions).

CHALLENGES FOR SECONDARY DATA ANALYSES

The use of the method of secondary data analysis has the following clear advantages for social researchers (Rew et al. 2000:226):

- It allows analyses of social processes in other inaccessible settings.
- It saves time and money.
- It allows the researcher to avoid data collection problems.
- It facilitates comparison with other samples.
- It may allow inclusion of many more variables and a more diverse sample than otherwise would be feasible.
- It may allow data from multiple studies to be combined.

The greatest challenge faced in secondary data analysis results from the researcher's inability to design data collection methods that are best suited to answer their research question. The secondary data analyst also cannot test and refine the methods to be used based on preliminary feedback from the population or processes to be studied. Nor is it possible for the secondary data analyst to engage in the iterative process of making observations, developing concepts, or making more observations and refining the concepts. This last problem is a special challenge for those seeking to conduct secondary analyses of qualitative data because an inductive process of developing research questions and refining observation and interview strategies is a hallmark of much qualitative methodology (Heaton 2008:511).

These limitations mean it may not be possible for a secondary data analyst to focus on the specific research question of original interest or to use the most appropriate sampling or measurement approach for studying that research question. Secondary data analysis inevitably involves a trade-off between the ease with which the research process can be initiated and the specific hypotheses that can be tested and methods that can be used. If the primary study was not designed to measure adequately a concept that is critical to the secondary analyst's hypothesis, the study may have to be abandoned until a more adequate source of data can be found. Alternatively, hypotheses, or even the research question itself, may be modified to match the analytic possibilities presented by the available data (Riedel 2000:53).

Data quality is always a concern with secondary data, even when the data are collected by an official government agency. Government actions result, at least in part, from political processes that may not have as their priority the design or maintenance of high-quality data for social scientific analysis. For example, political opposition to the British Census's approach to recording ethnic origin led to changes in the 1991 Census that rendered its results inconsistent with prior years and that demonstrated the "tenuous relationship between enumeration [Census] categories and possible social realities" (Fenton 1996:155).

It makes sense to use official records to study the treatment of juveniles accused of illegal acts because these records document the critical decisions to arrest, to convict, or to release (Dannefer and Schutt 1982). But research based on official records can be only as good as the records themselves. In contrast to the controlled interview process in a research study, there is little guarantee that the officials' acts and decisions were recorded in a careful and unbiased manner. The same is true for data collected by employees of private and nonprofit organizations. For example, research on the quality of hospital records has created, at best, mixed support for the validity of the key information they contain (Iezzoni 1997:391).

This one example certainly does not question all legal records or all other types of official records. It does, however, highlight the value of using multiple methods, particularly when the primary method of data collection is analysis of records generated by **street-level bureaucrats**—officials who serve clients and have a high degree of discretion (Lipsky 1980). When officials make decisions and record the bases for their decisions without much supervision, records may diverge considerably from the decisions they are supposed to reflect. More generally, it is always important to learn how people make sense of the social world when we want to describe their circumstances and explain their behavior (Needleman 1981). The secondary data analyst should review carefully the process by which data were collected and note the limitations discovered in research reports.

The basis for concern is much greater in research across national boundaries because different data collection systems and definitions of key variables may have been used (Glover 1996). Census counts can be distorted by incorrect answers to census questions as well as by inadequate coverage of the entire population (Rives and Serow 1988:32–35). National differences in the division of labor between genders within households can confuse the picture when comparing household earnings between nations without accounting for these differences (Jarvis 1997:521).

Reanalyzing qualitative data someone else collected also requires setting aside the expectation that qualitative research procedures and interpretations will be informed by intimate familiarity with the context in which the data were collected and with those from whom the data were obtained (Heaton 2008:511). Instead, the secondary analyst of qualitative data must seek opportunities for carrying on a dialogue with the original researchers.

Many of these problems can be lessened by seeking conscientiously to review data features and quality before deciding to develop an analysis of secondary data (Riedel 2000:55–69; Stewart and Kamins 1993:17–31) and then developing analysis plans that maximize the value of

the available data. Replicating key analyses with alternative indicators of key concepts, testing for the stability of relationships across theoretically meaningful subsets of the data, and examining findings of comparable studies conducted with other data sets can each strengthen confidence in the findings of a secondary analysis.

Any secondary analysis will improve if the analyst—yourself or the author of the work that you are reviewing—answers several questions before deciding to develop an analysis of secondary data in the first place and then continues to develop these answers as the analysis proceeds (adapted from Riedel 2000:55–69; Stewart and Kamins 1993:17–31):

1. *What were the agency's or researcher's goals in collecting the data?*

   The goals of the researcher, research, or research sponsor influence every step in the process of designing a research project, analyzing the resulting data, and reporting the results. Some of these goals will be stated quite explicitly, but others may only be implicit—reflected in the decisions made but not acknowledged in the research report or other publications. When you consider whether to use a data set for a secondary analysis, you should consider whether your own research goals are similar to those of the original investigator and sponsor. The data collected are more likely to include what is necessary for achieving your own research goals if the original investigator or sponsor had similar goals. When your research question or other goals diverge from those of the original investigator, you should consider how this divergence may have affected the course of the primary research project and whether this affects your ability to use the resulting data for a different purpose.

   For example, Pamela Paxton (2002) studied the role of secondary organizations in democratic politics in a sample of 101 countries but found that she could only measure the prevalence of international nongovernmental associations (INGOs) because comparable figures on purely national associations were not available. She cautioned, "INGOs represent only a specialized subset of all the associations present in a country" (Paxton 2002:261). We need to consider this limitation when interpreting the results of her secondary analysis.

2. *What data were collected, and what were they intended to measure?*

   You should develop a clear description of how data enter the data collection system, for what purpose, and how cases leave the system and why. Try to obtain the guidelines that agency personnel are supposed to follow in processing cases. Have there been any changes in these procedures during the period of investigation (Riedel 2000:57–64)?

3. *When was the information collected?*

   Both historical and comparative analyses (see Chapter 15) can be affected. For example, the percentage of the U.S. population not counted in the U.S. Census appears to have declined since 1880, from about 7% to 1%, but undercounting continues to be more common among poorer urban dwellers and recent immigrants (King and Magnuson 1995; see also Chapter 5). The relatively successful 2000 U.S. Census reduced undercounting (Forero 2000b) but still suffered from accusations of shoddy data collection procedures in some areas (Forero 2000a).

4. *What methods were used for data collection? Who was responsible for data collection, and what were their qualifications?* Are they available to answer questions about the data?

   Each step in the data collection process should be charted and the involved personnel identified. The need for concern is much greater in research across national

boundaries because different data collection systems and definitions of key variables may have been used (Glover 1996). Incorrect answers to census questions as well as inadequate coverage of the entire population can distort census counts (seeChapter 5; Rives and Serow 1988:32–35). Copies of the forms used for data collection should be obtained, specific measures should be inspected, and the ways in which these data are processed by the agency or agencies should be reviewed.

5. *How is the information organized (by date, event, etc.)?* Are there identifiers that are used to identify the different types of data available (Riedel 2000:58–61)?

   Answers to this set of questions can have a major bearing on the work that will be needed to carry out the study.

6. *What is known about the success of the data collection effort? How are missing data indicated? What kind of documentation is available? How consistent are the data with data available from other sources?*

   The U.S. Census Bureau provides extensive documentation about data quality, including missing data, and it documents the efforts it makes to improve data quality. The 2020 Census 0Program for Evaluations, Assessments, and Experiments was designed to improve the decennial census in 2020, as well as other Census Bureau censuses and surveys (Hill and Reichert 2021). This has been an ongoing effort, since 1950, with tests of questionnaire design and other issues. You can read more about it at https://www.census.gov/newsroom/blogs/random-samplings/2021/09/2020-census-program-for-evaluations-experiments-and-assessments.html.

Answering these six questions helps ensure that the researcher is familiar with the data he or she will analyze and can help identify any problems with it. It is unlikely that you or any secondary data analyst will be able to develop complete answers to all these questions before starting an analysis, but it still is critical to attempt to assess what you know and don't know about data quality before deciding whether to conduct the analysis. If you uncover bases for real concern after checking documents, other publications based on the data, information on websites, and perhaps by making some phone calls, you may have to decide to reject the analytic plan and instead search for another data set. If your initial answers to these six questions give sufficient evidence that the data can reasonably be used to answer your research question, you should still keep seeking to fill in missing gaps in your initial answers to the questions; through this ongoing process, you will develop the fullest possible understanding of the quality of your data. This understanding can lead you to steer your analysis in the most productive directions and can help you write a convincing description of the data set's advantages and limitations.

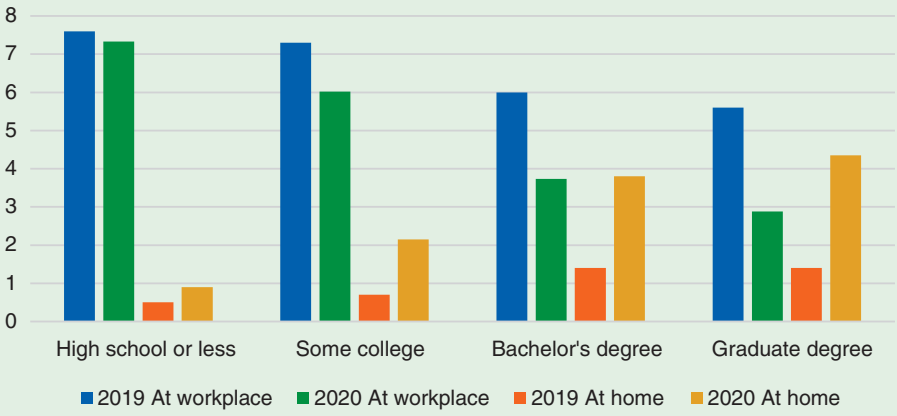
RESEARCH IN THE NEWS

THE PANDEMIC CHANGED HOW WE SPENT OUR TIME

The Department of Labor's Bureau of Labor Statistics conducts the American Time Use Survey annually. Thousands of people track in their survey responses every minute of their time in a single day. By comparing time use from the year before the pandemic to the same period in 2020, BLS researchers learned a lot about the impact of the pandemic on daily lives. Time spent alone rose by almost one hour, while time spent with people outside the household

declined by 1 ½ hours. Changes in work location differed dramatically by education level (see Figure 14.7). While workers with no more than a high school education continued to spend almost all their work time at the workplace, those with more education—particularly those with a graduate degree—shifted a major portion of their work time to their home.

FIGURE 14.7 ■ Time Spent at Workplace and at Home, May–December 2019–2020



Source: Casselman and Koeze. 2021:B6. Data from American Time Use Survey, Bureau of Labor Statistics.

For Further Thought?

1. What additional variables would you like to consider in an analysis that tries to clarify the association between education level and the likelihood of shifting to working at home?
2. What new research question would you like to focus on in an analysis of the 2019–2020 American Time Use Survey data?

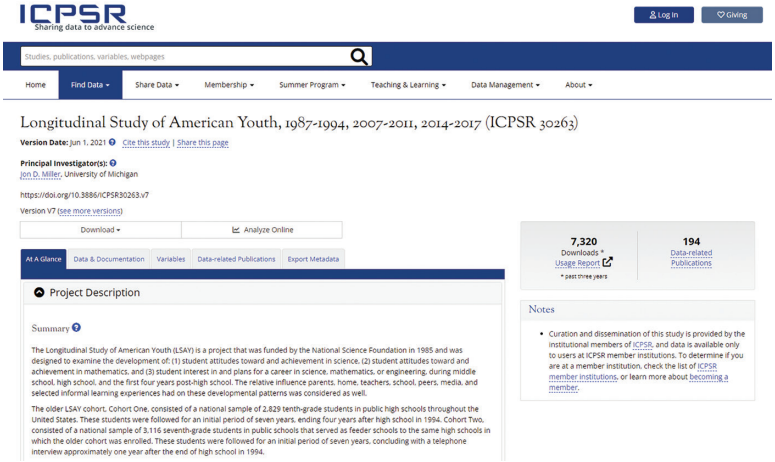
Source: Casselman, Ben, and Ella Koeze. 2021. "The Pandemic Changed How We Spent Our Time." *The New York Times*, July 29:B1, B6.

This seems like a lot to ask, doesn't it? After all, you can be married for life after answering only one question; here, I'm encouraging you to attempt to answer six questions before committing yourself to a brief relationship with a data set. Fortunately, the task is not normally so daunting. If you acquire a data set for analysis from a trusted source, many of these questions will already have been answered for you. You may need to do no more than read through a description of data available on a website to answer the secondary data questions and consider yourself prepared to use the data for your own purposes. If you are going to be conducting major analyses of a data set, you should take more time to read the complete study documents, review other publications with the data, and learn about the researchers who collected the data.

Figure 14.8 contains the description of longitudinal survey data about American youth available from ICPSR. Read through it and see how many of the secondary data questions it answers.

In an environment in which so many important social science data sets are quickly available for reanalysis, the method of secondary data analysis should permit increasingly rapid refinement of social science knowledge, with new hypotheses tested and methodological disputes clarified. Both the necessary technology and the supportive ideologies required for this rapid refinement have spread throughout the world. Social science researchers now can take advantage of this methodology but must also accept the responsibility to delineate and acknowledge the method's limitations carefully and publicly.

FIGURE 14.8 ■ ICPSR Data Set Description


ICPSR
Sharing data to advance science
Log in
Giving
Studies, publications, variables, webpages
Home
Find Data
Share Data
Membership
Summer Program
Teaching & Learning
Data Management
About
Longitudinal Study of American Youth, 1987-1994, 2007-2011, 2014-2017 (ICPSR 30263)
Version Date: Jun 1, 2021 Cite this study | Share this page
Principal Investigator(s):
Jon D. Miller, University of Michigan
https://doi.org/10.3886/ICPSR30263.v7
Version V7 (see more versions)
Download
Analyze Online
At A Glance
Data & Documentation
Variables
Data-related Publications
Export Metadata
Project Description
Summary
The Longitudinal Study of American Youth (LSAY) is a project that was funded by the National Science Foundation in 1985 and was designed to examine the development of: (1) student attitudes toward and achievement in science, (2) student attitudes toward and achievement in mathematics, and (3) student interest in and plans for a career in science, mathematics, or engineering, during middle school, high school, and the first four years post-high school. The relative influence parents, home, teachers, school, peers, media, and selected informal learning experiences had on these developmental patterns was considered as well.
The older LSAY cohort, Cohort One, consisted of a national sample of 2,829 tenth-grade students in public high schools throughout the United States. These students were followed for an initial period of seven years, ending four years after high school in 1994. Cohort Two, consisted of a national sample of 3,116 seventh-grade students in public schools that served as feeder schools to the same high schools in which the older cohort was enrolled. These students were followed for an initial period of seven years, concluding with a telephone interview approximately one year after the end of high school in 1994.
7,320
Downloads *
Usage Report
* past three years
194
Data-related Publications
Notes
• Curation and dissemination of this study is provided by the institutional members of ICPSR, and data is available only to users at ICPSR member institutions. To determine if you are at a member institution, check the list of ICPSR member institutions, or learn more about becoming a member.

Source: Inter-university Consortium for Political and Social Research.

BIG DATA

"We have come to expect information to appear when and where we need it, and to be able to communicate with anyone at any time, no matter where they are" (Abernathy 2017:2). Do you agree? The smartphone in your hand, pocket, or purse, with its text messaging, e-mail, video chat, and other capabilities; the GPS in your car or on your phone; the social media sites you visit; the information available to you through Google searches—these all tie you to a boundless world of information from and about people. And all of these connections leave digital tracks—"bread crumbs"—that provide data about social behavior (Abernathy 2017:54), ranging from pictures and videos to tweets and text (Lazer and Radford 2017). **Big Data** analyses are now being used to predict the spread of flu, the price of airline tickets, the behavior of consumers—and to investigate the social world.

Background

Computers began to connect people in new ways in 1969, when computers at the University of California, Los Angeles, were linked to computers at the Stanford Research Institute (and two other universities) through a network known as ARPANET—after its funder, the U.S. Defense Department's Advanced Research Projects Agency. The network, which later became known as the Internet, expanded for the next two decades with enhancements in e-mail, programming and computer hardware and the involvement of more organizations. In 1990, British computer scientist Sir Tim Berners-Lee developed the first web browser, making it much easier for users to manage their computer-mediated connections. The use of the Internet began to explode after the World Wide Web became publically available in 1991. During the same period, the U.S. Air Force created a global positioning system (GPS) by launching 24 special satellites in evenly spaced orbits that covered the globe by 1995 (now 27 are in active operation at any one time). These GPS satellites allowed two-way communication of location with units on the ground—units that now include your smartphone. The "geoweb" was born, and vast amounts of interconnected data were available as never before (Abernathy 2017:19–26). And, of course, interpersonal interactions were changed too, as spatial media "increasingly mediate social interactions within spaces and provide different ways to know and navigate locales" (Kitchin, Lauriault, and Wilson 2017:11).

The quantity of data generated in this interconnected web is astounding. In December 2021, there were over 5 billion Internet users (almost two thirds of the world's population), more than 3 billion active Facebook users, and 1.91 million websites, with more than 200 billion e-mails sent daily (https://www.internetlivestats.com/). Every minute, people post more than 4 million "likes" in Facebook and send more than 340,000 tweets (for more examples, see Figure 14.9). Each day, there are more than 6 billion Google searches and about that many videos viewed on YouTube, and also more than 70 million photos uploaded in Instagram (http://www.internetlivestats.com/). That's "Big."

FIGURE 14.9 ■ Activity in One Internet Minute

In One Minute Every Day Internet Users

- Make 110,040 Skype calls
- Ride Uber 694 times
- Like content on Facebook 4,166,667 times
- Tweet 347,222 times
- Upload 300 hours of video on YouTube
- Like Instagram photos 1,736,111 times
- Pin photos on Pinterest 9,722 times
- Download 51,100 apps from Apple
- Stream 77,160 hours of video on Netflix
- Vote 18,327 times on Reddit
- Visit Amazon 4,310 times (unique visitors)
- Play 1,041,666 videos on Vine
- Swipe 590,278 times on Tinder
- Share snaps 284,722 on SnapChat
- View 34,130 videos on Buzzfeed

Source: https://bgr.com/tech/internet-activity-one-minute/; Data from Domo.com

The sources of Big Data are increasing rapidly. More than 3 billion people use Facebook, thereby creating digital records that can, with appropriate arrangements, be analyzed to better understand social behavior (http://www.internetlivestats.com/) (Aiden and Michel 2013:12; Desjardins 2017). Big Data are also generated by GPS users, social media, smartphones, wristband health monitors, student postings, and even student activity in online education programs (Mayer-Schönberger and Cukier 2013:90–96, 115). A new Big Data system records preemies' heart rates, respiration rates, and temperatures—what amounts to 1,260 data points per second—and can predict the onset of infection 24 hours before the appearance of overt symptoms (Mayer-Schönberger and Cukier 2013:60). Public utilities, government agencies, and private companies can all learn about their customers from analyzing patterns revealed in their records. Although much of these data are inaccessible to those who are not given permission by the organizations that collect the data and none of it is of value for those who lack the appropriate computing power and statistical skills, it is all creating possibilities for investigations of the social world that could not have been envisioned even 20 years ago. Rather than relying on self-reports of behaviors such as social contacts, the

digital traces of these behaviors can be studied directly through smartphone data on physical proximity (given participants' consent), data on Facebook contacts, and many other types of digitized data (Lazer and Radford 2017).

Examples of Research Using Big Data

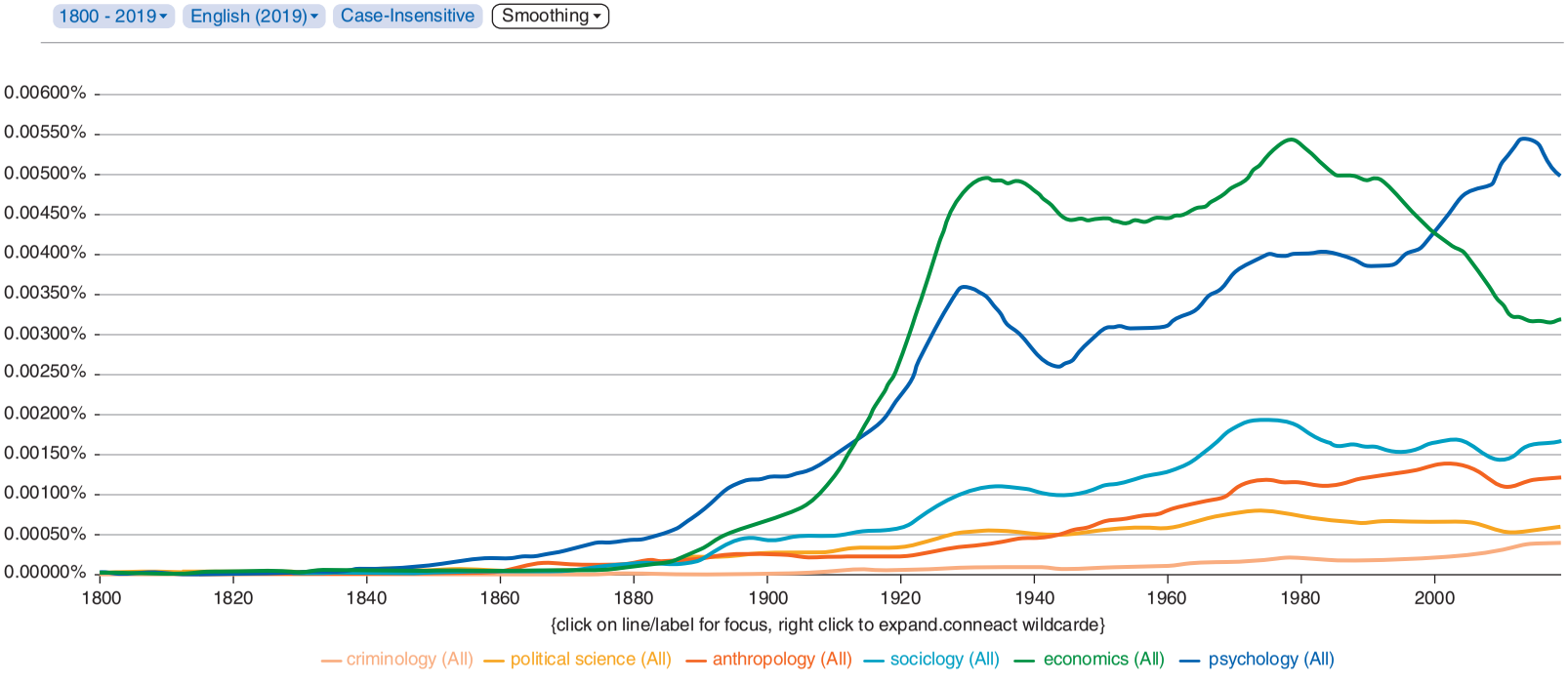
How can you use such data in social science research? Here's a quick example. What could be more important than an interest in sociology and other social sciences? After all, if you have declared sociology or another social science as your major, you probably think that this discipline is having a positive impact in the social world; maybe you want to contribute something to that impact yourself. So, would you like to know how popular your discipline is? One way to answer that question is to see how frequently the name of the discipline has appeared in all the books ever written in the world. It may surprise you to learn that it is possible right now to answer that question, although with two key limitations: We can only examine books written in English and in several other languages, and as of 2014, we are limited to "only" one quarter of all books ever published—30 million books (Aiden and Michel 2013:16).

To check this out, go to the Google **Ngrams** site (https://books.google.com/ngrams), type in "sociology, political science, anthropology, criminology, psychology, economics," and check the "case-insensitive" box (and change the ending year to 2022). Figure 14.10 shows the resulting screen (if you don't obtain a graph, try using a different browser). Note that the height of a graph line represents the percentage that the term represents of all words in books published in each year, so a rising line means greater relative interest in the word, not simply more books being published. You can see that psychology emerges in the mid-19th century, while sociology, economics, anthropology, and political science appear in the latter part of that century, and criminology arrives in the early 20th century. You can see that interest in sociology increased as the 1960s progressed, but then dropped off in the 1980s. What else can you see in the graph? It's hard to stop checking other ideas by adding in other terms, searching in other languages, or shifting to another topic entirely. Now, that's not the same as checking how many people are reading these words in these books, but it still provides quite an idea of what authors have written about. (For other limitations of Ngrams, see Zhang 2015.)

Dan O'Brien at Northeastern University and Chris Winship at Harvard (O'Brien and Winship 2017) used geographically based Big Data to investigate the importance of place in fluctuations in crime. They calculated indexes of violent crime, physical disorder, and social disorder at specific addresses from over 2 million geocoded emergency and non-emergency calls to the city of Boston's 911 (emergency) and 311 numbers (non-emergency reports of disorder, such as potholes or graffiti). Fewer than 1% of addresses generated 25% of reports of crime and disorder, and almost all variation in crime was between addresses. In other words, "problem properties" were responsible for most crime and disorder in Boston. You can check out the data resources and mapping capabilities that O'Brien and Winship used at the Boston Area Research Initiative (BARI) website (https://www.northeastern.edu/csshresearch/bostonarearesearchinitiative/).

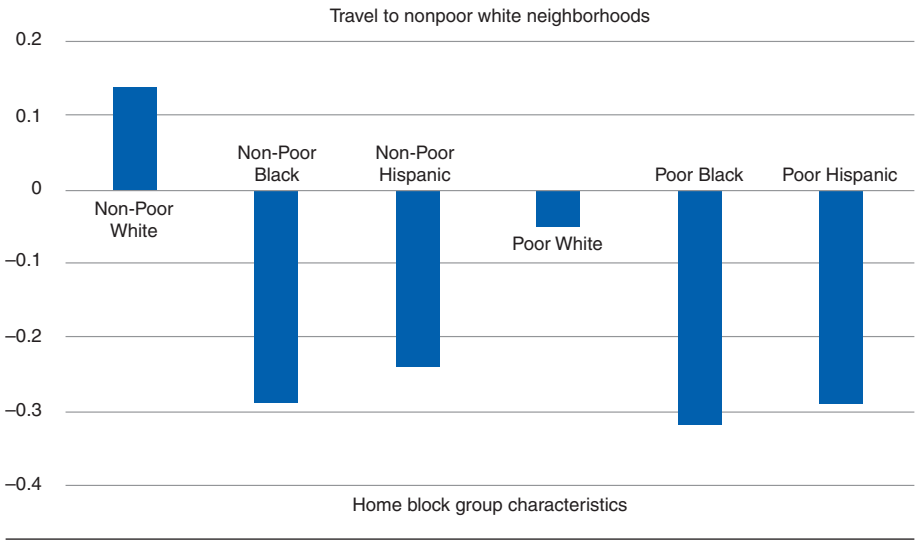
An analysis of mobility patterns in America's 50 largest cities got even "bigger." Qi Wang, Nolan Edward Phillips, Mario L. Small, and Robert J. Sampson analyzed 650 million geocoded Twitter messages to estimate the home locations and travel patterns of about 40,000 residents during an 18-month period (Wang et al. 2018). They found that mobility patterns tended to sustain social isolation, even though residents of disadvantaged neighborhoods tended to travel as much as those from advantaged neighborhoods. Residents of primarily Black and primarily Hispanic neighborhoodshad much less exposure to nonpoor or white neighborhoods than residents of primarily white neighborhoods, regardless of income level, while white poor neighborhoods were substantially isolated from non-poor white neighborhoods (see Figure 14.11).

FIGURE 14.10 ■ Ngram of Social Sciences



Source: Google Books Ngram Viewer, https://books.google.com/ngrams. Accessed December 21, 2021.

FIGURE 14.11 ■ Proportions of Individuals Traveling to Non-poor White Neighborhoods, by Block Group Characteristics



Source: Wang, Qi, Nolan Edward Phillips, Mario L. Small, and Robert J. Simpson. 2018. "Urban Mobility and Neighborhood Isolation in America's 50 Largest Cities." *PNAS* 115(30):7735-7740.

In a widely discussed but cautionary tale about an attempt to use Big Data to improve public health, Jeremy Ginsberg and some colleagues at Google set out to speed up the response to the spread of flu by tracking the online searches of the 90 million U.S. adults who seek information about specific illnesses each year (Ginsberg et al. 2009:1012). By comparing the Google search data with official data collected by the U.S. Centers for Disease Control and Prevention (CDC), Ginsberg and his colleagues were able to identify search trends indicating when people first start to experience symptoms (Butler 2013:155).

But the cautionary tale is this: In the 2013 flu season, Google Flu Trends predicted a much higher peak level of flu than occurred. The problem seems to have been that widespread media coverage and the declaration of a public health emergency in New York led many more people than usual to search for flu-related information, even though they were not experiencing symptoms themselves. Google has been refining its procedures to account for this problem, while other researchers have shifted their attention to analysis of flu-related "tweets" or to data from networks of thousands of volunteers who report symptoms experienced by family members to a central database (Butler 2013). So having incredible amounts of data does not allow us to forget about the potential for problems in measurement or sampling.

Digital Methods

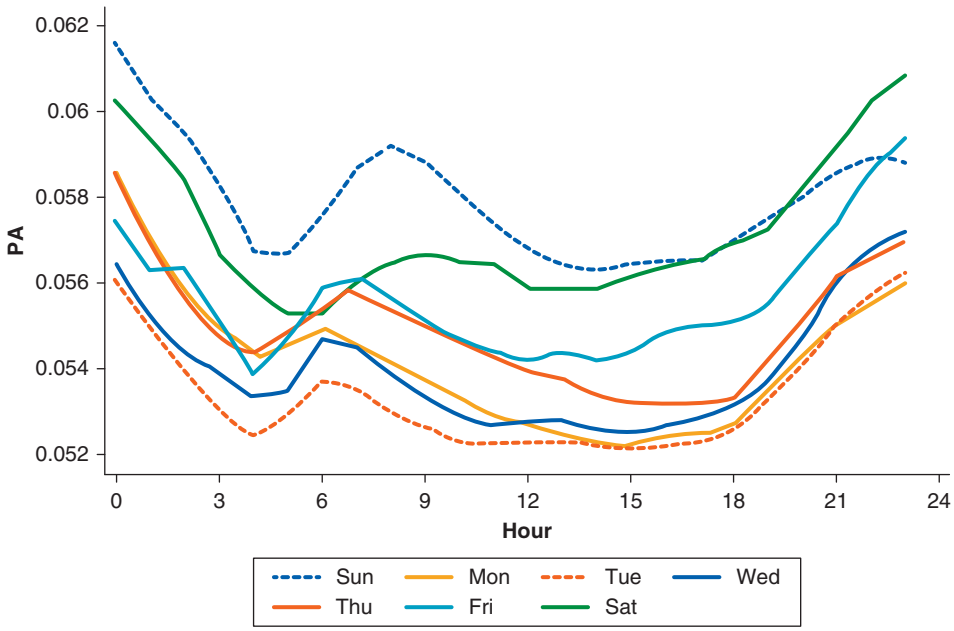
As substantial portions of social interaction continue to migrate online, particularly after the pandemic reduced much in-person social interaction, analyzing this online social activity has become increasingly important to understand its role in the social world. This is different than when online Big Data is analyzed as a record of human activities in the physical world, such as in the preceding examples of studying Google Ngrams (reflecting the contents of written books), BARI's area-based records on crime (based on 911 calls) and on mobility (based on physical locations of smartphones), and Google Flu Trends (responses to physical symptoms). However, an important part of the social world now begins and ends with online activity. Tweets, Facebook

posts, Google searches, and Wikipedia entries are themselves forms of social activity that tell us about the social world. **Digital methods** use online tools and online data to conduct research on online activity and how it is shaped by the features of the online medium. The patterns of use of digital objects such as hyperlinks, hashtags, timestamps, likes, shares, and retweets allow us to chart social activity and to determine how the use of digital objects is shaped by the structuring of these objects. They are strategies for dealing with the ephemeral and unstable nature of online data (Rogers 2019:3).

The study by Yulin Hswen and her team of Twitter hashtags and anti-Asian sentiment with which I began this chapter is a good example of the use of digital methods. It raises questions for further exploration with digital methods about how hashtags are used, whether sentiment expressed on Twitter can be generalized to the population, and whether analyses of activity on other social media platforms such as Facebook would show the same patterns (Hswen et al. 2021:962).

Two more examples of using Big Data show how digital methods can investigate important online social patterns. Sociologists Scott Golder and Michael Macy (2011) investigated mood fluctuations throughout the day and across the globe solely by analyzing 509 million messages posted on Twitter by 2.4 million individuals in 84 countries in 2008 and 2009. Using a standard system for identifying words expressing positive and negative affect (such feelings as "anxiousness," "anger," and "inhibition"), they found a common pattern of people awakening in a good mood that deteriorates throughout the day (see Figure 14.12). Intrigued? You too can search Tweets at https://twitter.com/search-advanced.

FIGURE 14.12 ■ Hourly Changes in Individual Positive Affect (PA) by Day of the Week



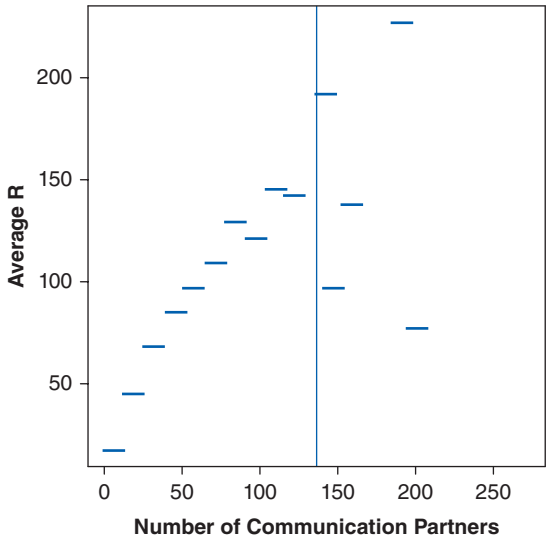
Source: Golder, Scott A., and Michael W. Macy. 2011. "Diurnal and Seasonal Mood Vary With Work, Sleep, and Daylength Across Diverse Cultures." *Science* 333:1878. Reprinted with permission from AAAS.

Angela Bohn, Christian Buchta, Kurt Hornik, and Patrick Mair, in Austria and at Harvard in the United States explored the relation between friendship patterns and access to social capital based on records from 438,851 Facebook users (Bohn et al. 2014). Bohn et al. (2014:32) started

their analysis with data on 1,712 users (they didn't have the computer power to analyze more data) who were selected randomly over a two-month study period, from about 1.3 million users who had agreed on Facebook to have their data used anonymously for such a study.

Figure 14.13 displays one of their findings about social networks: Having more communication partners increased social capital—as indicated by responses received to their postings—up to about 130 partners. Many Facebook users with more partners than that tended to receive fewer responses to their postings (Bohn et al. 2014:39). Having more partners can lead to too much of what otherwise is a good thing!

FIGURE 14.13 ■ Social Capital and Number of Communication Partners on Facebook



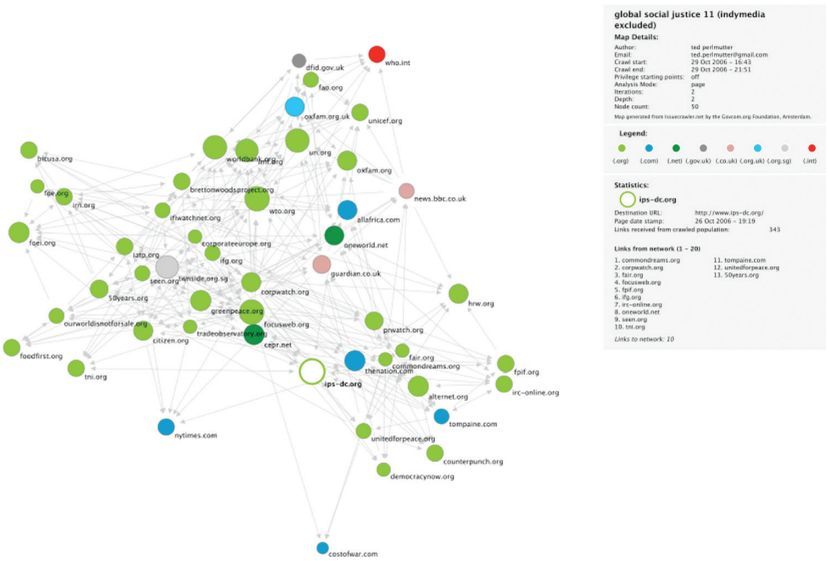
Source: Bohn, Angela, Christian Buchta, Kurt Hornik, and Patrick Mair. 2014. "Making Friends and Communicating on Facebook: Implications for the Access to Social Capital." *Social Networks* 37:29–41.

Using digital methods requires attention to new methodological issues. How should search queries be designed to capture online social activity about an issue? For example, in the Israeli–Palestinian conflict, the phrases "security fence" and "separation barrier" both refer to the same physical object, but they reflect different interpretations of the reasons for that physical object and indicate opposing political positions in the conflict. Studies of Google searches or tweets about the conflict would have to include both terms (and others) and could use them to investigate changes in sentiment (Rogers 2019:28-31). In another use of digital methods, Issuecrawler software may be used to locate networks of organizations engaged in the same issue area and chart the connections between them (Rogers 2019:51-55) (see Figure 14.14).

Challenges for Using Big Data

Although the number of cases in Big Data analyses can be enormous, they often do not represent a population that can be defined clearly in relation to generalizations (Lazer and Radford 2017). Digital data may be unrepresentative of the general population due to socioeconomic differences between those who use smartphones or connect to the Internet in other ways and those who live mostly "offline," as well as between data sets to which we are allowed access and

FIGURE 14.14 ■ Global Social Justice Network, Top 50 Actors, Issuecrawler.net Cluster Map Output, October 2006



Source: Richard Rogers, http://www.govcom.org/publications/full_list/rogers_issuecrawler_context.pdf. Accessed December 22, 2021.

those that are controlled by private companies (Lewis 2015). Which slice of the social world do Twitter users represent? Facebook subscribers? Instagram users? These concerns make it important to review research on the types of persons who use the platform(s) from which Big Data is being collected. Using data from multiple sources to attempt to replicate findings will also help in assessing the generalizability of conclusions from a particular source. In addition, the Big Data analyst also must consider how people may use multiple platforms for the same purpose—for example, connecting to people through smartphones, Zoom, and Google Hangouts—and then note the limitations of using only one or even a few sources to study a research question. Social behavior online may also not reflect behavior in the everyday world (Golder and Macy 2014:141–144).

It can even be a mistake to assume that Big Data directly represent the decisions and actions of individuals (Lazer and Radford 2017). Platforms like Google can change their operational procedures in ways that influence the likelihood of particular search results appearing. Automated bots may increase the number of "likes" or other indicators of interaction with sites, and people may use fraudulent or multiple identities when engaging on the web. *Caveat emptor!*

ETHICAL ISSUES WITH SECONDARY DATA AND BIG DATA

Analysis of data collected by others does not create the same potential for immediate harm as does the collection of primary data, but neither ethical nor related political considerations can be ignored. Big Data creates new ethical concerns, in part because it can reflect the behavior of many individuals who did not consent to participate in research.

Subject confidentiality is a key concern when original records are analyzed with either secondary data or Big Data. Whenever possible, all information that could identify individuals should be removed from the records to be analyzed so that no link is possible to the identities of living subjects or the living descendants of subjects (Huston and Naylor 1996:1698). When

you use data that have already been archived, you need to find out what procedures were used to preserve subject confidentiality. The work required to ensure subject confidentiality probably will have been done for you by the data archivist. For example, ICPSR examines carefully all data deposited in the archive for the possibility of disclosure risk. All data that might be used to identify respondents are altered to ensure confidentiality, including the removal of information such as birth dates or service dates, specific incomes, or place of residence that could be used to identify subjects indirectly (see http://www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted/index.html). If all information that could be used in any way to identify respondents cannot be removed from a data set without diminishing data set quality (e.g., by preventing links to other essential data records), ICPSR restricts access to the data and requires that investigators agree to conditions of use that preserve subject confidentiality. Those who violate confidentiality may be subject to a scientific misconduct investigation by their home institution at the request of ICPSR (Johnson and Bullock 2009:218). The UK Data Archive provides more information about confidentiality and other human subjects protection issues at https://www.ukdataservice.ac.uk/manage-data/legal-ethical. The problem of identification can be compounded when using Big Data, since Internet service providers can track users in their activities and may maintain identifying information. Ethics boards may require users of Big Data to limit their analyses to the specific research proposed and, if personal identification may put persons at risk, require informed consent from those whose data is included in an analysis (Weinhardt 2020).

It is not up to you to decide whether there are any issues of concern regarding human subjects when you acquire a data set for secondary analysis from a responsible source. The institutional review board (IRB) for the protection of human subjects at your college or university or other institution has the responsibility to decide whether they need to review and approve proposals for secondary data analysis. For example, the IRB at the University of Wisconsin–Madison (UW–Madison) waives review for specific secondary data sets that meet human subjects protection standards and allows researchers to request exemption from review for other data sources that are demonstrated to meet these standards (https://kb.wisc.edu/gradsch/page.php?id=29465). Specifically, their regulations stipulate that research projects involving secondary data set analysis will *not* require prior IRB approval, if the data set has been preapproved by the UW–Madison IRB as indicated by posting on a list that includes the following data sets:

- ICPSR
- University of Wisconsin Data and Information Services Center (DISC)
- Roper Center for Public Opinion Research
- U.S. Census Bureau
- National Center for Health Statistics
- National Center for Education Statistics
- National Election Studies

Data sets that may qualify for inclusion on UW–Madison's list of approved data sources include the following:

- Public use data sets posted on the Internet that include a responsible use statement or other confidentiality agreement for authors to protect human subjects (e.g., see ICPSR's responsible use statement)

- Survey data distributed by UW principal investigators who can certify that (1) the data collection procedures were approved by a qualified IRB meeting the Common Rule criteria for IRB and (2) the data set and documentation as distributed do not contain information that could be used to identify individual research participants
- *Note:* Research projects that merge more than one data set in such a way that individuals may be identified are not covered by this policy and require prior IRB approval.

Political concerns intersect with ethical practice in secondary data analyses. How are *race* and *ethnicity* coded in the U.S. Census? You learned in Chapter 4 that changing conceptualizations of race have affected what questions are asked in the census to measure race. This data collection process reflects, in part, the influence of political interest groups, and it means that analysts using the census data must understand why the proportion of individuals choosing "other" as their race and the proportion in a "multiracial" category has changed. The same types of issues influence census and other government statistics collected in other countries. Britain's census first asked about ethnic group in 1991. British researcher Steve Fenton (1996) reports that the design of the specific questions and categories used to measure ethnic groups "was clearly based on a conception of ethnic minorities as constituted by Black Caribbean and Asian populations" (p. 156). Respondents were asked to classify themselves only as white or Black (in several subcategories), Indian, Pakistani, Bangladeshi, Chinese, or "any other ethnic group."

Other concerns can be much greater in research across national boundaries because different data collection systems and definitions of key variables may have been used (Glover 1996). Government funding decisions can affect the availability of government statistics on particular social issues (Levitas and Guy 1996:3). Census counts can be distorted by inadequate coverage of the entire population (see Chapter 5; Rives and Serow:198832–35). Social and political pressures may influence the success of a census in different ways in different countries. Some Mexicans were concerned that the results of Mexico's 2000 Census would be "used against them" by the government, and nearly 200,000 communities were inaccessible for follow-up except by a day's mule travel (Burke 2000). In rural China, many families that had flouted the government's official one-child policy sought to hide their "extra" children from census workers (Rosenthal 2000).

Because in most cases the secondary researchers did not collect the data, a key ethical obligation is to cite the original, principal investigators, as well as the data source, such as ICPSR. Researchers who seek access to data sets available through Europe's CESSDA must often submit a request to the national data protection authority in the country (or countries) of interest (Johnson and Bullock 2009:214).

Big Data also creates some new concerns about research ethics. When enormous amounts of data are available for analysis, the usual procedures for making data anonymous may no longer ensure that it stays that way (Weinhardt 2020). In 2006, AOL released for research purposes 20 million search queries from 657,000 users, after all personal information had been erased and only a unique numeric identifier remained to link searches. However, staff at *The New York Times* conducted analyses of sets of search queries and were able quickly to identify a specific individual user by name and location, based on that user's searches. The collection of Big Data also makes possible surveillance and prediction of behavior on a large scale. Crime control efforts and screening for terrorists now often involve developing predictions from patterns identified in Big Data. Without strict rules and close monitoring, potential invasions of privacy and unwarranted suspicions are enormous (Mayer-Schönberger and Cukier 2013, pp. 150–163). Should researchers be able to analyze tweets without the consent of the "tweeters" (Moe and Larsson 2012)? Consulting the latest ethical guidelines from organizations like

the Association of Internet Researchers will help to chart a reasonable approach (Weinhardt 2020:350).

Combining different types of Big Data and data from different platforms can also create risks for those whose data is obtained (Weinhardt 2020). Linking biographical data and "likes" from Facebook, political statements from blogs or tweets, contacts collected by smartphone apps, and pictures from Instagram could make it easy to identify political dissidents or individuals who violate conventional norms in their country of residence. The inherent cross-national and longitudinal feature of data available on the web heightens risks that individuals could in some circumstances be identified. The inability to obtain informed consent from individuals represented in massive datasets heightens the need for researchers to avoid putting unknowing participants in their research at heightened risk. When data added on the web is open to the public, such as tweets and blogs, informed consent is not an issue, but if users have to log in to a platform, they may consider their actions within that platform to be available only to those they designate or others on that platform.

Social experiments with Big Data can literally change the social environment, and so this too raises ethical concerns. In a striking example, Robert Bond, James Fowler, and others at Facebook and the University of California, San Diego, conducted a randomized experiment with Facebook on the day of the 2010 congressional elections (Bond et al. 2012:295). Here is their description of the research design:

> Users [of Facebook] were randomly assigned to a "social message" group, an "informational message" group, or a control group. The social message group (n = 60,055,176) was shown a statement at the top of their "News Feed." This message encouraged the user to vote, provided a link to find local polling places, showed a clickable button reading "I Voted," showed a counter indicating how many other Facebook users had previously reported voting, and displayed up to six small randomly selected "profile pictures" of the user's Facebook friends who had already clicked the I Voted button. . . . The informational message group (n = 611,044) was shown the message, poll information, counter, and button, but they were not shown any faces of friends. The control group (n = 613,096) did not receive any message at the top of their News Feed.

As indicated in Figure 14.15, individuals in the group that received the personalized message about their friends having voted were more likely to vote—and the effect was higher the more closely connected they were to those friends. Bond et al. (2012:297) estimate that receiving this message brought 60,000 more voters to the polls in 2010, whereas the effect of friends having seen the message could have increased turnout by 280,000 votes!

Do you believe that researchers should be able to conduct experiments that may alter the behavior of thousands of people? What if it changes the result of an election? Is it okay if the behavior that is increased by the experiment is widely accepted as a public good—like increasing voter turnout? How do these issues compare to those you considered in Chapter 7 about recruiting introductory psychology students for an experiment in a small group lab.

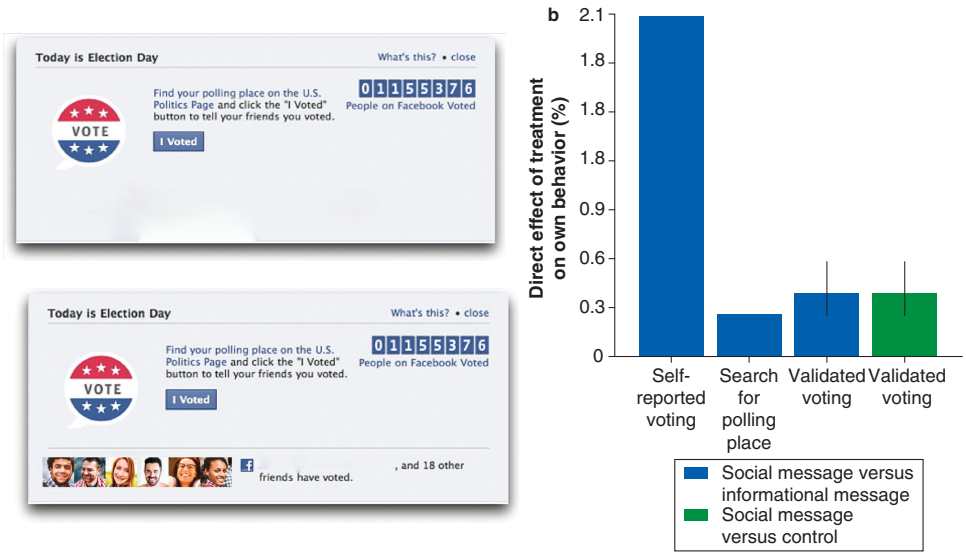
Michael Weinhardt (2020) at the Institute of Sociology, Technische Universität Berlin, suggests adhering to several ethical guidelines for Big Data research:

- Follow existing formal requirements (including laws and user agreements).
- State clearly the extent to which a research project is about identifiable human subjects.
- Use data from "proper" sources that are documented and maintain standards of privacy.

- Assess possible risks for those represented in the data and conduct a careful risk-benefit analysis—concerning all potentially affected parties and the severity and likelihood of the different risks.
- Take account particular standards and rules in the specific jurisdictions from which data are collected.

And be sure to review your procedures with local experts in Big Data research.

FIGURE 14.15 ■ Facebook Experiment on Voter Turnout



Examples of (a) the informational message and social message Facebook treatments and (b) their direct effect on voting behavior. Vertical lines indicate s.e.m. (they are too small to be seen for the first two bars).

Sources: Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489:295–298.

CONCLUSIONS

The easy availability for secondary analyses of data sets collected in thousands of social science investigations is one of the most exciting features of social science research in the 21st century. You can often find a previously collected data set that is suitable for testing new hypotheses or exploring new issues of interest. Moreover, the research infrastructure that has developed at ICPSR and other research consortia, both in the United States and internationally, ensures that a great many of these data sets have been carefully checked for quality and archived in a form that allows easy access. The availability of Big Data records that reflect behavior both offline and online creates entirely new possibilities for investigating the social world, although quality checks mostly remain up to the researcher.

Many social scientists now review available secondary data and Big Data possibilities before they consider collecting new data with which to investigate a particular research question. Even

if you do not leave this course with a plan to become a social scientist yourself, you should now have the knowledge and skills required to find and use secondary data and to review analyses of Big Data to answer your own questions about the social world.

KEY TERMS

Big Data (p. 505)
Digital methods (p. 510)
Inter-university Consortium for Political and Social Research (ICPSR) (p. 491)
Ngrams (p. 507)
Secondary data (p. 490)
Secondary data analysis (p. 490)
Street-level bureaucrats (p. 501)

HIGHLIGHTS

LO 14.1 Explain the steps involved in using ICPSR.

- ICPSR provides the most comprehensive social science data archive for academic researchers in the United States.
- Begin a search in ICPSR by identifying keywords and phrases that capture your research focus and/or specific variables that would be used in your analysis.
- Review the study descriptions, publications, and variables associated with each possible dataset.
- If there is more than one dataset that seems to match your needs, consider downloading more than one dataset and exploring the data in more depth.
- Record a proper citation for the dataset.
- Search SocINDEX or another bibliographic database to learn about related studies that used different databases.

LO 14.2 Describe at least four major sources of secondary data for social scientists.

- The U.S. Bureau of the Census provides official population statistics based on the decennial census, with much data reported for geographic units.
- The Integrated Public Use Microdata Series (IPUMS) at the University of Minnesota's Minnesota Population Center provides samples from the decennial U.S. Census and its annual American Community Survey, as well as data collected by censuses conducted in more than 100 other countries. Survey data are also included on topics like health and education and the environment.
- The U.S. Department of Labor's Bureau of Labor Statistics (BLS) provides data on employment, earnings, prices, living conditions, industrial relations, productivity and technology, and occupational safety and health, collected from surveys of individuals and of establishments. BLS also sponsors the National Longitudinal Surveys (NLS) of children, youth, and adults.
- Other government sources of secondary data are the National Technical Information Service (NTIS) of the U.S. Department of Commerce and state governments.
- The Add Health study funded by the National Institute of Child Health and Human Development (NICHD) and 23 other agencies and foundations provides data on more than 20,000 persons first interviewed as adolescents.

- The Health and Retirement Study (HRS), funded by the National Institute on Aging, provides ongoing data on more than 22,000 Americans over the age of 50. Other family-oriented secondary datasets are the National Survey of Families and Households and the Fragile Families & Child Wellbeing Study.
- The Roper Center archives public opinion poll data, with more than 23,000 U.S. and international data sets.
- Harvard University's Dataverse Project at its Institute for Quantitative Social Science includes more than 50,000 datasets in the social sciences.
- Sources of international data include the U.S. Social Security Administration's data on social security programs, the European Commission's Eurobarometer Survey Series, and the Council of European Social Science Data Archives.
- Qualitative data is available from the UK Data Service QualiBank, from the ICPSR, and from the Harvard Dataverse in the United States, as well as from the Human Relations Area Files at Yale University.

LO 14.3 Defend the use of secondary data, considering the challenges it poses.

- Secondary data analysts should have a good understanding of the research methods used to collect the data they analyze. Data quality is always a concern.
- Available secondary data may not match the desired research question or specific hypotheses, but the research questions or hypotheses may be modified to create a closer fit.
- The secondary data analyst should carefully review the process by which data were collected and note the limitations discovered in research reports.
- Inconsistencies in data collection or variable definitions are potential problems when using official data from international sources and must be reviewed carefully.
- Secondary analysts of qualitative data may seek to develop a dialogue with the original researchers to compensate for their inability to probe the original participants in more depth.
- Replicating key analyses with alternative indicators of key concepts, testing for the stability of relationships across theoretically meaningful subsets of the data, and examining findings of comparable studies conducted with other data sets can each strengthen confidence in the findings of a secondary analysis.
- Before and during a secondary data analysis, determine the original data collection goals, the specific measures and other methods used, the dates data were collected, and the ways in which the data are organized and documented.

LO 14.4 Assess the strengths and weaknesses of research using Big Data.

- Collection of massive sets of Big Data permits analysis of large-scale social patterns and trends.
- Big Data sources have proliferated, and many are readily accessible.
- Many sources provide direct measures of behavior that do not depend on self-report.
- Individuals may change their behavior in response to the release of Big Data analyses.

- Differing policies of social media and other platforms can shape the data generated.
- Despite large numbers of records, Big Data may not represent a clearly definable population of interest.
- Search queries must be carefully designed to capture all relevant data.

LO 14.5 Plan an IRB presentation about the ethics of a Big Data research project.

- Most secondary datasets available from ICPSR and other public sources have had identifying information removed.
- Preserving subject confidentiality is a key concern with Big Data, so methods to preserve confidentiality when the data were first collected should be reviewed and procedures adopted to avoid retaining information that could identify individual respondents.
- Review local IRB requirements for working with secondary data and take these into account in designing human subjects protection procedures.
- Consider the expectations of those using sites to which access is restricted before concluding that data should be treated as publicly available.
- Be particularly careful to avoid linking data from different sources that could increase the likelihood of personal identification and lead to disclosures that put individuals in some countries or communities at risk.
- Consult the latest ethnical guidelines from professional associations.
- Social experiments with Big Data can literally change the social environment, and so this too raises ethical concerns that should be considered.

DISCUSSION QUESTIONS

1. What are the strengths and weaknesses of secondary data analysis? Do you think it's best to encourage researchers to try to address their research questions with secondary data if possible?
2. What are the similarities and differences between secondary data analysis and Big Data analysis? Do you feel one of these approaches is more likely to yield valid conclusions? Explain your answer.
3. In a world of limited resources and time constraints, should social researchers be required to include in their proposals to collect new data an explanation of why they cannot investigate their proposed research question with secondary data? Such a requirement might include a systematic review of data that already are available at ICPSR and other sources. Discuss the merits and demerits of such a requirement. If such a requirement were to be adopted, what specific rules would you recommend?
4. What is your experience with differences in the type of users and the motivations for use of different online platforms? How do recommend the differences in platforms like Facebook and Twitter be taken into account in plans for analysis of research questions using Big Data?

SPSS EXERCISE

This is the time to let your social scientific imagination run wild because any analysis of the GSS data set will qualify as a secondary data analysis! Review the list of variables in the GSS dataset you've been using, formulate one hypothesis involving two of these variables *that have no more than five categories*, and test the hypotheses using the cross-tabulation procedure. (Specify the independent variable as the column variable in the selection window, specify the dependent variable as the row variable, and then choose the Option for your table of column percents.) See if the percentage distribution of the dependent variable varies across the categories of the independent variable in either or both tables and propose an explanation for your findings.

©iStockphoto.com/hermosawave

15 HISTORICAL AND COMPARATIVE METHODS AND CONTENT ANALYSIS

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

15.1 Illustrate the value of historical and comparative methods with one historical or contemporary example

15.2 Contrast the two types of historical social science methods

15.3 Contrast the three types of comparative social science methods

15.4 Appraise the value of adding demographic analysis to historical and comparative investigations

15.5 Critique content analysis as a way of identifying social patterns

15.6 Formulate an argument for exempting historical and comparative research from IRB review

Although the United States and several European nations have maintained democratic systems of governance for more than 100 years, democratic rule has more often been brief and unstable, when it has occurred at all. In the last decade, more authoritarian regimes have emerged, and democratic rule has been threatened in many countries, including within the United States and Europe. What explains the maintenance of democratic practices in one country and their absence in another? Are democratic politics a realistic option for every nation? Are there some prerequisites in historical experience, cultural values, or economic resources (Markoff 2005:384–386)? Is democracy at risk even in the United States?

Historical and comparative research methods help to answer questions like these with a diverse set of methodological tools that allow investigation of social processes at other times and in other places. These methods involve several different approaches and a diverse set of techniques, and they may have qualitative or quantitative components. They provide ways to investigate topics that usually cannot be studied with experiments, participant observation, or surveys. Because this broader focus involves collecting data from records about the past or from other nations, the methods used in historical and comparative investigations present unique challenges.

In this chapter, I will review the major methods social scientists use to understand historical processes and to compare different societies or regions. I also will introduce oral histories, a qualitative tool for historical investigations, as well as demographic methods, which can strengthen both historical and comparative studies. In addition, we will study the method of content analysis, which can be used in historical and comparative research as well as in contemporary studies of communication with any type of media. Throughout the chapter, I will draw many examples from research on democracy and the process of democratization.

OVERVIEW OF HISTORICAL AND COMPARATIVE RESEARCH METHODS

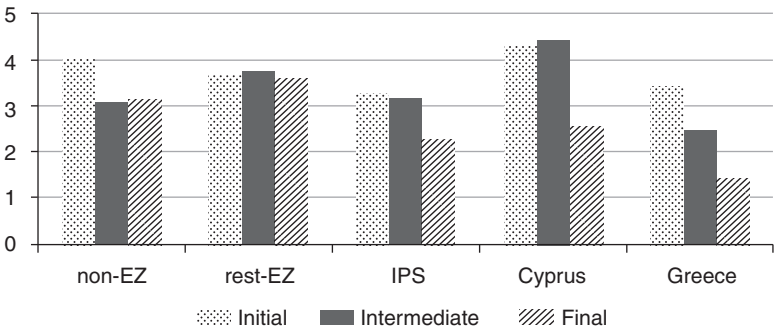
The central insight behind historical and comparative research is that we can improve our understanding of social processes when we make comparisons to other times and places. Max Weber's comparative study of world religions (Bendix 1962) and Émile Durkheim's (1893/2014)

historical analysis of the division of labor are two examples of the central role of historical and comparative research during the period sociology emerged as a discipline. Although the popularity of this style of research ebbed with the growth of survey methods and statistical analysis in the 1930s, exemplary works, such as Reinhard Bendix's (1956) *Work and Authority in Industry* and Barrington Moore Jr.'s (1966) *Social Origins of Democracy and Dictatorship,* helped to fuel a resurgence of historical and comparative methods in the 1970s and 1980s, which has continued into the 21st century (Lange 2013:22–33).

Research by José Luís Casanova (2018) at Portugal's Instituto Universitario de Lisboa about the authoritarian trend in Europe demonstrates how useful historical and comparative methods are for answering important current research questions about the course of democracy. After the 2008 global financial crisis, several countries in the European Union struggled to meet their financial obligations and sought financial assistance from international institutions: the European Commission, the European Central Bank, and the International Monetary Fund (IMF) (Hall and Reis 2015). The resulting bailout packages came with requirements for an adjustment program limiting public spending to allow reimbursement of the loans but creating greater financial hardships for the population of these countries.

Did these austerity programs lead to changes in political culture in these states? To answer this research question, Casanova compared changes in trust in government and other indicators of political culture measured in the European Social Survey between countries that had had adjustment programs (Greece and Cyprus and Ireland, Portugal, and Spain [IPS]) and other Eurozone (EZ) countries (that use the Euro as their currency) and the non-Eurozone countries in the European Union. You can see in Figure 15.1 that the countries subject to an adjustment program (IPS, Cyprus, Greece) experienced a sharp drop in trust in politicians (and on many other indicators) after the adaptation program was in place.

FIGURE 15.1 ■ Trust in Politicians by Eurozone Membership and Adaptation Program



Initial = 2002; Intermediate = 2008; Final = 2014

Source: José Luís Casanova. 2018. "Crisis and Cultural Change: The Countries With Adjustment Programmes in the European Union." *Comparative Sociology* 17:187–220 (p. 192). ©Brill, Brill.com.

While Casanova's (2018) research demonstrates the value of historical and comparative methods, these methods include a more diverse collection of approaches that can involve combinations of other methods presented in this text (Lange 2013). Research may be historical, comparative, or both historical and comparative (like Casanova's). There are no hard-and-fast rules for determining how far in the past the focus of research must be to consider it historical

or what types of comparisons are needed to warrant calling research comparative. In practice, research tends to be considered historical when it focuses on a period before the experience of most of those conducting research (Abbott 1994:80). Research involving different nations is usually considered comparative, but so are studies of different regions within one nation if they emphasize interregional comparison. In recent years, the globalization of U.S. economic ties and the internationalization of scholarship have increased the use of comparative research methods across many different countries (Kotkin 2002). Historical and comparative methods can be quantitative or qualitative or a mixture of both. Both nomothetic and idiographic approaches to establishing causal effects can be used.

Distinguishing research with a historical or comparative focus results in four basic types of research: **historical events research**, **historical process research**, **cross-sectional comparative research**, and **comparative historical research**. Research that focuses on events in one short historical period is historical events research, whereas longitudinal research that traces a sequence of events over a number of years is historical process research (see, for example, Skocpol 1984:359). There are also two types of comparative research, the first involving cross-sectional comparisons and the second comparing longitudinal data about historical processes between multiple cases. The resulting four types of research are displayed in Figure 15.2.

FIGURE 15.2 ■ Types of Historical and Comparative Research

	Cross-Sectional	Longitudinal
Single Case	Historical Events Research	Historical Process Research
Multiple Cases	Cross-Sectional Comparative Research	Comparative Historical Research

HISTORICAL SOCIAL SCIENCE METHODS

Both historical events research and historical process research investigate questions concerning past times. These methods are used increasingly by social scientists in sociology, anthropology, political science, and economics, as well as by many historians (Monkkonen 1994). The late 20th and early 21st centuries have seen so much change in so many countries that many scholars have felt a need to investigate the background of these changes and to refine their methods of investigation (Hallinan 1997; Robertson 1993). The accumulation of large bodies of data about the past not only has stimulated more historically oriented research but also has led to the development of several different methodologies.

Much historical (and comparative) research is qualitative. This style of historical social science research tends to have several features that are similar to those used in other qualitative methodologies. First, like other qualitative methods, qualitative historical research is inductive:

It develops an explanation for what happened from the details discovered about the past. In addition, qualitative historical research is **case oriented**; it focuses on the nation or other unit as a whole rather than only on different parts of the whole in isolation from each other (Ragin 2000:68). This could be considered the most distinctive feature of qualitative research on historical processes. The research question is "What was Britain like at the time?" rather than "What did Queen Elizabeth do?" Related to this case orientation, qualitative historical research is **holistic**—concerned with the context in which events occurred and the interrelations between different events and processes: "how different conditions or parts fit together" (Ragin 1987:25–26). For the same reason, qualitative historical research is **conjunctural** because, it is argued, "no cause ever acts except in complex conjunctions with others" (Abbott 1994:101). Charles Ragin (2000:67–68) uses the example of case-oriented research on the changing relationship between income and single parenthood in the United States after World War II:

> In the end, the study is also about the United States in the second half of the twentieth century, not just the many individuals and families included in the analysis. More than likely, the explanation of the changing relation between income and single parenthood would focus on interrelated aspects of the United States over this period. For example, to explain the weakening link between low income and single parenthood the researcher might cite the changing status of women, the decline in the social significance of conventional family forms, the increase in divorce, the decrease in men's job security, and other changes occurring in the United States over this period.

Qualitative historical research is also **temporal** because it looks at the related series of events that unfold over time. It is therefore also likely to be *historically specific*—limited to the specific time(s) and place(s) studied. Qualitative historical research uses **narrative explanations**—idiographic causal reasoning (see Chapter 6)—in which the research tells a story involving specific actors and other events occurring at the same time (Abbott 1994:102) or one that accounts for the position of actors and events in time and in a unique historical context (Griffin 1992). Larry Griffin's (1993) research on lynching, in the next section, provides a good example.

The focus on the past presents special methodological challenges:

- Documents and other evidence may have been lost or damaged.
- Available evidence may represent a sample biased toward more newsworthy figures.
- Written records will be biased toward those who were more prone to writing.
- Feelings of individuals involved in past events may be hard, if not impossible, to reconstruct.

Before you judge historical social science research as credible, you should look for convincing evidence that each of these challenges has been addressed.

Historical Events Research

Research on past events that does not follow processes for some long period is historical events research rather than historical process research. Historical events research basically uses a cross-sectional rather than longitudinal design. Investigations of past events may be motivated by the belief that they had a critical impact on subsequent developments or because they provide opportunities for testing the implications of a general theory (Kohn 1987).

Event-Structure Analysis

One technique useful in historical events research, as well as in other types of historical and comparative research, is **event-structure analysis**. Event-structure analysis is a qualitative approach that relies on a systematic coding of key events or national characteristics to identify the underlying structure of action in a chronology of events. The codes are then used to construct event sequences, make comparisons between cases, and develop an idiographic causal explanation for a key event.

An event-structure analysis consists of the following steps:

1. Classifying historical information into discrete events
2. Ordering events into a temporal sequence
3. Identifying prior steps that are prerequisites for subsequent events
4. Representing connections between events in a diagram
5. Eliminating from the diagram connections that are not necessary to explain the focal event

Griffin (1993) used event-structure analysis to explain a unique historical event, a lynching in the 1930s in Mississippi. According to published accounts and legal records, the lynching occurred after David Harris, an African American who sold moonshine from his home, was accused of killing a white tenant farmer after an argument about payment became heated. After the killing was reported, a local deputy sheriff was called and a citizen search party was formed, without any investigation or trial. The deputy did not intervene as the search party trailed Harris and then captured and killed him. Meanwhile, Harris's friends killed another African American who had revealed Harris's hiding place. This series of events is outlined in Figure 15.3.

–––Which among the numerous events occurring between the time that the tenant farmer confronted Harris and the time that the mob killed Harris had a causal influence on that outcome? To identify these idiographic causal links, Griffin identified plausible counterfactual possibilities—events that might have occurred but did not—and considered whether the outcome might have been changed if a counterfactual had occurred instead of a particular event (see Chapter 6):

> If, contrary to what actually happened, the deputy had attempted to stop the mob, might the lynching have been averted? . . . Given what happened in comparable cases and the Bolivar County deputy's clear knowledge of the existence of the mob and of its early activities, his forceful intervention to prevent the lynching thus appears an objective possibility. (Griffin 1993:1112)

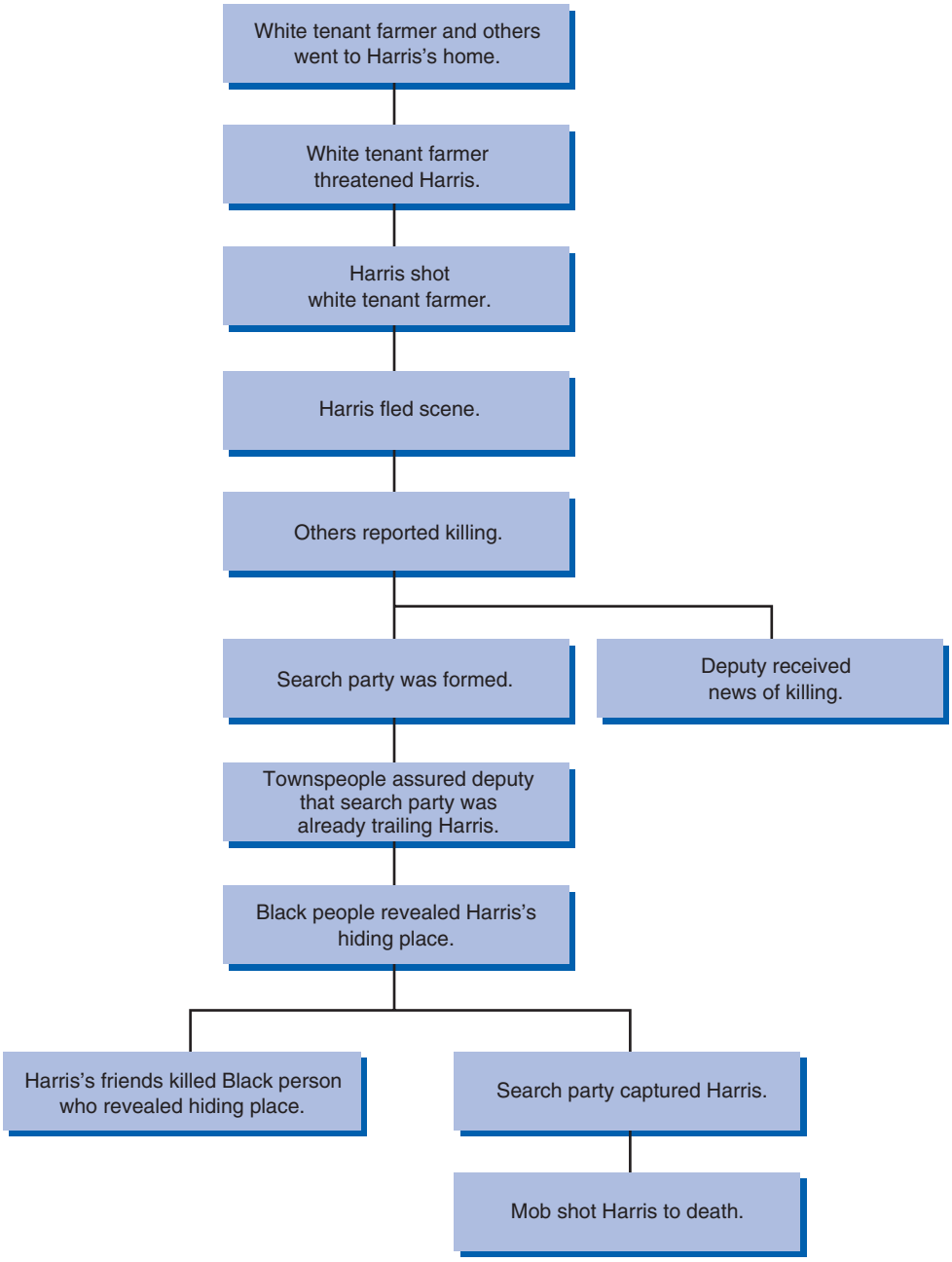
So Griffin concluded that nonintervention by the deputy had a causal influence on the lynching.

Oral History

History that is not written down is mostly lost to posterity (and to social researchers). However, **oral histories** can be useful for understanding historical events that occurred within the lifetimes of living individuals. As the next example shows, sometimes oral histories even result in a written record that can be analyzed by researchers at a later point in time.

Thanks to a Depression-era writers' project, Deanna Pagnini and Philip Morgan (1996) found that they could use oral histories to study attitudes toward out-of-wedlock births among African American and white women in the South during the 1930s.

FIGURE 15.3 ■ Event-Structure Analysis: Lynching Incident in the 1930s


White tenant farmer and others went to Harris's home.
White tenant farmer threatened Harris.
Harris shot white tenant farmer.
Harris fled scene.
Others reported killing.
Search party was formed.
Deputy received news of killing.
Townspeople assured deputy that search party was already trailing Harris.
Black people revealed Harris's hiding place.
Harris's friends killed Black person who revealed hiding place.
Search party captured Harris.
Mob shot Harris to death.

Source: Adapted from Griffin (1993:1110).

Almost 70% of African American babies are born to unmarried mothers, compared with 22% of white babies (Pagnini and Morgan 1996:1696). This difference often is attributed to contemporary welfare policies or problems in the inner city, but Pagnini and Morgan thought it might be due to more enduring racial differences in marriage and childbearing. The history of slavery heightens the importance of investigating such explanations. To investigate these historical differences, Pagnini and Morgan read 1,170 life histories recorded by almost 200 writers who

worked for a New Deal program during the Depression of the 1930s, the Federal Writers' Project southern life history program. The interviewers had used a topic outline that included family issues, education, income, occupation, religion, medical needs, and diet.

In 1936, the divergence in rates of nonmarital births was substantial in North Carolina: 2.6% of white births were to unmarried women, compared with 28.3% of nonwhite births. The oral histories gave some qualitative insight into community norms that were associated with these patterns. A white seamstress who became pregnant at age 16 recalled, "I'm afraid he didn't want much to marry me, but my mother's threats brought him around" (Pagnini and Morgan 1996:1705). There were some reports of suicides by unwed young white women who were pregnant. In comparison, African American women who became pregnant before they were married reported regrets but rarely shame or disgrace. There were no instances of young Black women dying by suicide or getting abortions in these circumstances.

> We found that bearing a child outside a marital relationship was clearly not the stigmatizing event for African Americans that it was for whites. . . . When we examine contemporary family patterns, it is important to remember that neither current marriage nor current childbearing patterns are "new" for either race. Our explanations for why African Americans and whites organize their families in different manners must take into account past behaviors and values. (Pagnini and Morgan 1996:1714–1715)

Whether oral histories are collected by the researcher or obtained from an earlier project, the stories they tell can be no more reliable than the memories that are recalled. Unfortunately, memories of past attitudes are "notoriously subject to modifications over time" (Banks 1972:67), as are memories about past events, relationships, and actions. Use of corroborating data from documents or other sources should be used when possible to increase the credibility of descriptions based on oral histories.

Historical Process Research

Historical process research extends historical events research by focusing on a series of events that happened over a longer period. This longitudinal component allows for a much more complete understanding of historical developments than is often the case with historical events research, although it often uses techniques that are also used for research on historical events at one point in time, such as event history analysis and oral histories.

Historical process research can also use quantitative techniques. The units of analysis in quantitative analyses of historical processes are nations or larger entities, and researchers use a longitudinal design to identify changes over time. For example, David John Frank, Ann Hironaka, and Evan Schofer (2000) treated the entire world as their "case" for their deductive test of alternative explanations for the growth of national activities to protect the natural environment during the 20th century. Were environmental protection activities a response to environmental degradation and economic affluence within nations, as many had theorized? Or, instead, were they the result of a "top-down" process in which a new view of national responsibilities was spread by international organizations? Frank et al.'s measures of environmental protectionism included the number of national parks among all countries in the world and memberships in international environmental organizations; one of their indicators of global changes was the cumulative number of international agreements (see Table 15.1 for a list of some of their data sources).

Figure 15.4a charts the growth of environmental activities identified around the world. Compare the pattern in this figure with the pattern of growth in the number of international

TABLE 15.1 ■ Variables for Historical Analysis of Environmental Protectionism

Dependent Variables	Definition	Data Source(s)	Period of Analysis
National parks and protected areas	Annual cumulative numbers of parks per nation–state	IUCN (1990)	1900–1990
Country chapters of international environmental nongovernmental associations	Annual numbers of chapters per nation–state	Fried (1905–1911); League of Nations (1921, 1938); UIA (1948–1990)	1900–1988
Nation-state memberships in intergovernmental environmental organizations	Annual numbers of memberships per nation–state	Fried (1905–1911); League of Nations (1921, 1938); UIA (1948–1990)	1900–1984
Environmental impact assessment laws	Year of founding	Wood (1995)	1966–1992
National environmental ministries	Year of founding	Europa Year Book (1970–1995)	1970–1995

Source: David John Frank, Ann Hironaka, and Evan Schofer. 2000. "The Nation–State and the Natural Environment over the Twentieth Century." *American Sociological Review.* 65(1):96–116.

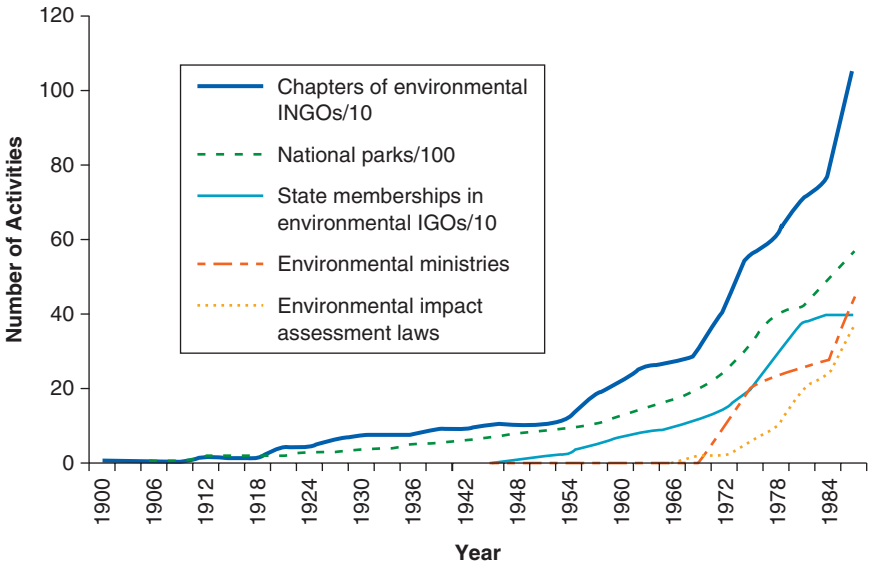
environmental agreements and national environmental laws shown in Figure 15.4b, and you can see that environmental protectionism at the national level was rising at the same time that it was becoming more the norm in international relations. In more detailed analyses, Frank and colleagues (2000) attempt to show that the growth in environmental protectionism was not explained by increasing environmental problems or economic affluence within nations. As in most research that relies on historical or comparative data, however, some variables that would indicate alternative influences (such as the strength of national environmental protest movements) could not be measured (Buttel 2000). Therefore, further research is needed.

Cautions for Historical Methods

One common measurement problem in historical research projects is the lack of data from some historical periods (Rueschemeyer, Stephens, and Stephens 1992:4; Walters, James, and McCammon 1997). For example, the widely used U.S. Uniform Crime Reporting System did not begin until 1930 (Rosen 1995). Sometimes, alternative sources of documents or estimates for missing quantitative data can fill in gaps (Zaret 1996), but even when measures can be created for key concepts, multiple measures of the same concepts are likely to be out of the question; as a result, tests of reliability and validity may not be feasible. Whatever the situation, researchers must assess the problem honestly and openly (Bollen, Entwisle, and Alderson 1993; Paxton 2002).

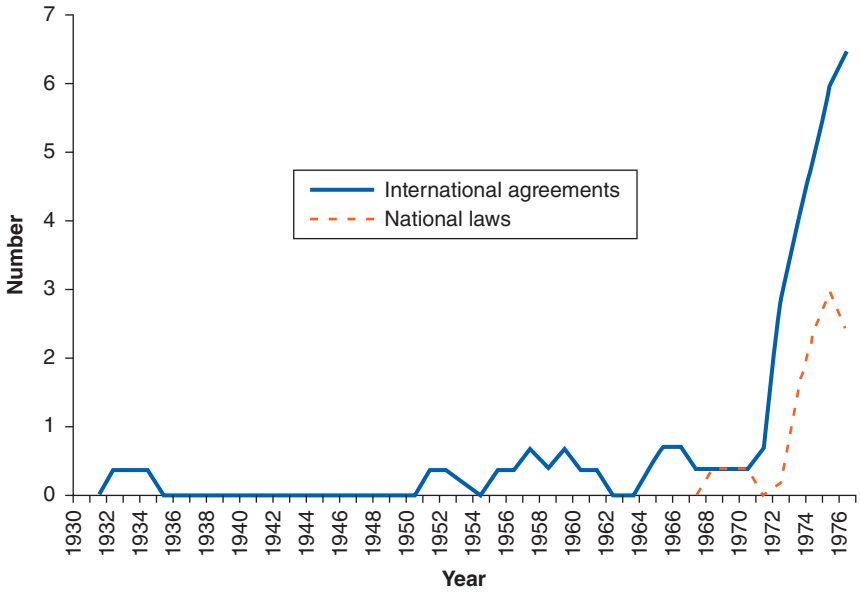
Those measures that are available are not always adequate. What is included in the historical archives may be an unrepresentative selection of materials that remain from the past. At various times, some documents could have been discarded, lost, or transferred elsewhere for a variety of reasons. *Original* documents may be transcriptions of spoken words or handwritten pages and could have been modified slightly in the process; they could also be outright distortions (Erikson 1966:172, 209–210; Zaret 1996). When relevant data are obtained from previous publications, it is easy to overlook problems of data quality, but this simply makes it all the more important to evaluate the primary sources. It is very important to develop a systematic plan for identifying and evaluating relevant documents.

FIGURE 15.4 ■ International Environmental Activity (a) Cumulative Numbers of Five National Environment Activities, 1900–1988 (b) Environmental Impact Assessment: International Agreements and National Laws, 1930–1977



(a) Cumulative Numbers of Five National Environment Activities, 1900–1988

Note: INGOs are international nongovernment organizations; IGOs are intergovernmental organizations.



(b) Environmental Impact Assessment: International Agreements and National Laws, 1930–1977

Source: David John Frank, Ann Hironaka, and Evan Schofer. February 2000. "The Nation-State and the Natural Environment over the Twentieth Century." *American Sociological Review*. 65(1):96–116.

COMPARATIVE SOCIAL SCIENCE METHODS

The limitations of single-case historical research have encouraged many social scientists to turn to comparisons between nations. These studies allow for a broader vision about social relations than is possible with cross-sectional research limited to one country or another unit. Both cross-sectional and longitudinal comparative methods encourage testing the cross-population generalizability of findings based on only one context (see Chapter 2).

Comparative Cross-Sectional Research

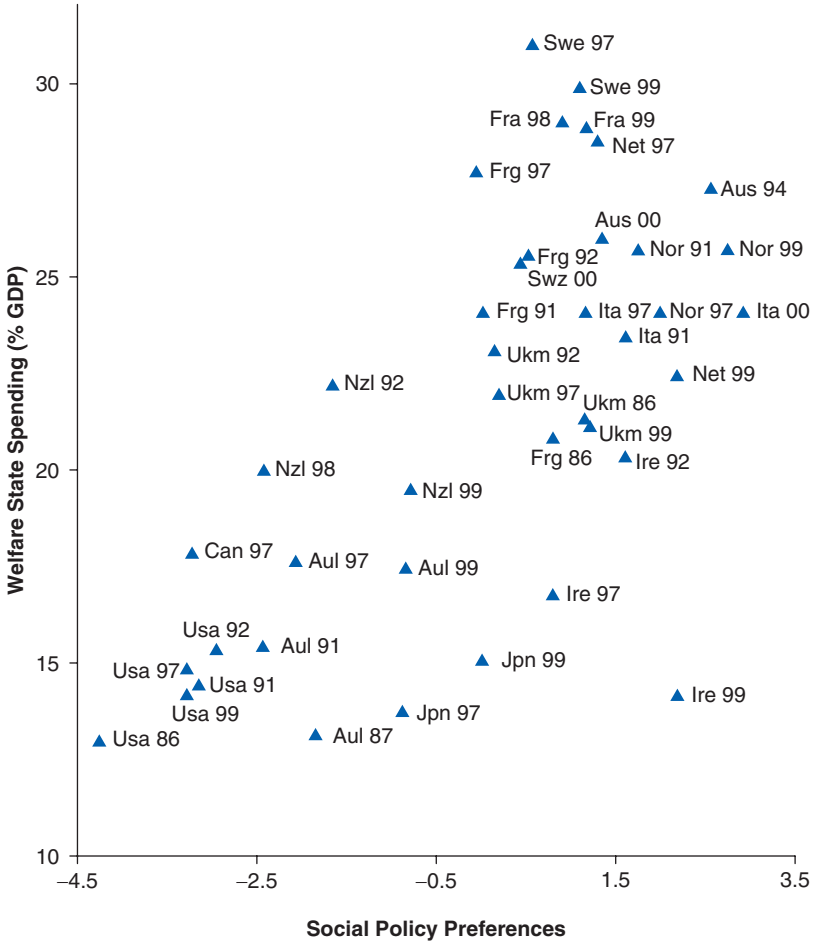
Comparisons between countries during one time period can help social scientists identify the limitations of explanations based on single-nation research. Such comparisons can suggest the relative importance of universal factors in explaining social phenomena compared with unique factors rooted in specific times and places (de Vaus 2008:251). These comparative studies may focus on a period in either the past or the present. Peter Houtzager and Arnab Acharya (2011) also point out that it can be more appropriate to compare cities or regions when the nations in which they are embedded vary internally in their social characteristics. For example, they compare the impact of engagement in associations on citizenship activity in São Paolo, Brazil, and Mexico City because the conditions exist for such an impact in these cities rather than in the surrounding countries.

Historical and comparative research that is quantitative may obtain data from national statistics or other sources of published data; if it is contemporary, such research may rely on cross-national surveys. Like other types of quantitative research, quantitative historical and comparative research can be termed **variable-oriented research**, with a focus on variables representing particular aspects of the units studied (Demos 1998).

Causal reasoning in quantitative historical and comparative research is nomothetic, and the approach is usually deductive, testing explicit hypotheses about relations between these variables (Kiser and Hechter 1991). For example, Clem Brooks and Jeff Manza (2006:476–479) deduce from three theories about welfare states—national values, power resources, and path dependency theory—the hypothesis that voters' social policy preferences will influence welfare state expenditures. Using country-level survey data collected by the International Social Survey Program (ISSP) in 15 democracies in five different years and expenditure data from the Organisation for Economic Co-operation and Development (OECD), Brooks and Manza were able to identify a consistent relationship between popular preferences for social welfare spending and the actual national expenditures (see Figure 15.5).

Popular preferences are also important factors in political debates over immigration policy. Christopher Bail (2008) asked whether majority groups in different European countries differ in the way that they construct "symbolic boundaries" that define "us" versus an immigrant "them." For his cross-sectional comparative investigation, he drew on 333,258 respondents in the 21-country European Social Survey (ESS). The key question about immigrants in the ESS was "Please tell me how important you think each of these things should be in deciding whether someone born, brought up and living outside [country] should be able to come and live here." The "things" whose importance they were asked to rate were six individual characteristics: (1) being white, (2) being well educated, (3) being from a Christian background, (4) speaking the official national language, (5) being committed to the country's way of life, and (6) having work skills needed in the country. Bail then calculated the average importance rating in each country for each of these characteristics and used a statistical procedure to cluster the countries by the extent to which their ratings and other characteristics were similar.

FIGURE 15.5 ■ Interrelationship of Policy Preferences and Welfare State Output

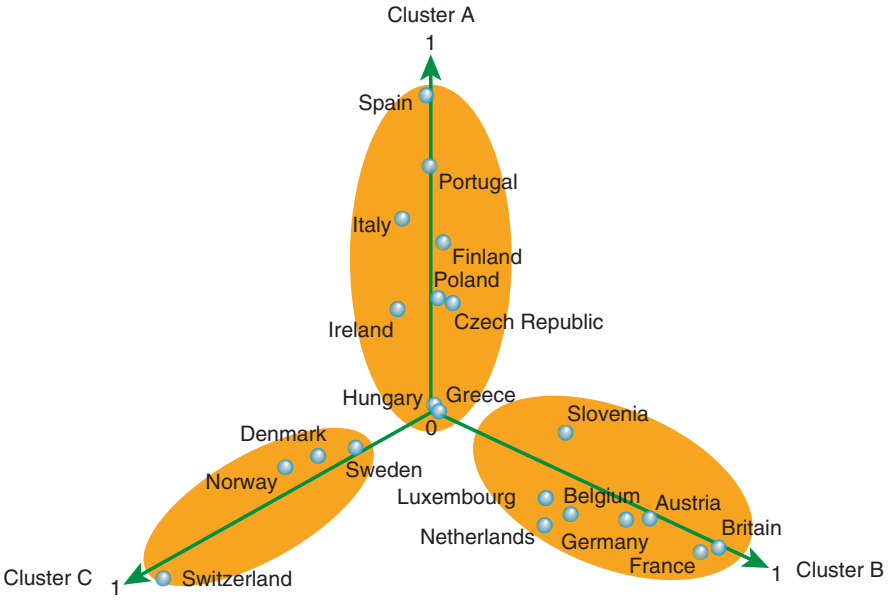


Note: Scattergram shows data for policy preferences and welfare state spending in 15 OECD democracies. Data are from the International Social Survey Program/Organisation for Economic Co-operation and Development.

Source: Brooks, Clem, and Jeff Manza. June 2006. "Social Policy Responsiveness in Developed Democracies." *American Sociological Review* 71(3):474–494.

Bail's (2008:54–56) analysis identified the countries as falling into three clusters (see Figure 15.6). Cluster A countries are on the periphery of Europe and have only recently experienced considerable immigration; their populations tend to draw boundaries by race and religion. Cluster B countries are in the core of Western Europe (except Slovenia); they have a sizable and long-standing immigrant population, and their populations tend to base their orientations toward immigrants on linguistic and cultural differences. All by one of the countries in Cluster C are in Scandinavia, have a varied but relatively large immigrant population, and attach much less importance to any of the six symbolic boundaries than do those in the other countries. Bail (2008:56) encourages longitudinal research to determine the extent to which these different symbolic boundaries are the product or the source of social inequality in these countries.

FIGURE 15.6 ■ Symbolic Boundaries Against Immigrants in 21 European Countries



Source: Bail, Christopher A. 2008. "The Configuration of Symbolic Boundaries Against Immigrants in Europe." *American Sociological Review* 73(1):37–59.

Cross-sectional comparative research has also helped explain variation in voter turnout. This research focuses on a critical issue in political science: Although free and competitive elections are a defining feature of democratic politics, elections themselves cannot orient governments to popular sentiment if citizens do not vote (LeDuc, Niemi, and Norris 1996). As a result, the low levels of voter participation in U.S. elections have long been a source of practical concern and research interest. International data give our first clue for explaining voter turnout: The historic rate of voter participation in the United States (48.3%, on average) is much lower than it is in many other countries that have free, competitive elections; for example, Italy has had a voter turnout of 92.5%, on average, since 1945 (see Table 15.2).

Does this variation result from differences between voters in knowledge and wealth? Do media and political party get-out-the-vote efforts matter? Mark Franklin's (1996:219–222) analysis of international voting data indicates that neither explanation accounts for much of the international variation in voter turnout. Instead, the structure of competition and the importance of issues are influential. Voter turnout is maximized where structural features maximize competition: compulsory voting (including, in Figure 15.6, Austria, Belgium, Australia, and Greece), mail and Sunday voting (including the Netherlands and Germany), and multiday voting. Voter turnout also tends to be higher where the issues being voted on are important and where results are decided by proportional representation (as in Italy and Israel, in Figure 15.6) rather than on a winner-take-all basis (as in U.S. presidential elections)—so individual votes are more important.

Franklin concludes that it is these characteristics that explain the low level of voter turnout in the United States rather than the characteristics of individual voters. The United States lacks the structural features that make voting easier, the proportional representation that increases the

TABLE 15.2 ■ Average Percentage of Voters Who Participated in Presidential or Parliamentary Elections, 1945–1998*

Country	Vote %	Country	Vote %
Italy	92.5	St. Kitts and Nevis	58.1
Cambodia	90.5	Morocco	57.6
Seychelles	96.1	Cameroon	56.3
Iceland	89.5	Paraguay	56.0
Indonesia	88.3	Bangladesh	56.0
New Zealand	86.2	Estonia	56.0
Uzbekistan	86.2	Gambia	55.8
Albania	85.3	Honduras	55.3
Austria	85.1	Russia	55.0
Belgium	84.9	Panama	53.4
Czech	84.8	Poland	52.3
Netherlands	84.8	Uganda	50.6
Australia	84.4	Antigua and Barbuda	50.2
Denmark	83.6	Burma/Myanmar	50.0
Sweden	83.5	Switzerland	49.3
Mauritius	82.8	USA	48.3
Portugal	82.4	Mexico	48.1
Mongolia	82.3	Peru	48.0
Tuvalu	81.9	Brazil	47.9
Western Samoa	81.9	Nigeria	47.6
Andorra	81.3	Thailand	47.4
Germany	80.9	Sierra Leone	46.8
Slovenia	80.6	Botswana	46.5
Aruba	80.4	Chile	45.9
Namibia	80.4	Senegal	45.6
Greece	80.3	Ecuador	44.7
Guyana	80.3	El Salvador	44.3
Israel	80.0	Haiti	42.9
Kuwait	79.6	Ghana	42.4
Norway	79.5	Pakistan	41.8
San Marino	79.1	Zambia	40.5
Finland	79.0	Burkina Faso	38.3

Country	Vote %	Country	Vote %
Suriname	77.7	Nauru	37.3
Malta	77.6	Yemen	36.8
Bulgaria	77.5	Colombia	36.2
Romania	77.2	Niger	35.6

*Based on entire voting-age population in countries that held at least two elections during these years. Only countries with highest and lowest averages are shown.

Source: Reproduced by permission of International IDEA from "Turnout in the World—Country by Country Performance (1945–1998)." From *Voter Turnout: A Global Survey* (http://www.int/vt/survey/voter_turnout_pop2-2.cfm) © International Institute for Democracy and Electoral Assistance.

impact of individuals' votes, and, often, the sharp differences between candidates that are found in countries with higher turnout. Because these structural factors are relatively stable within nations, we would never realize their importance if our analysis was limited to data from individuals in one nation. The increasing voter turnout in presidential elections as U.S. politics have become increasingly polarized also demonstrates the impact of the extent of perceived differences between the candidates (and the value of combining historical with comparative methods, the focus of the next section) (Romero and Romero 2021).

Despite the unique value of comparative analyses like Franklin's (1996), such cross-national research also confronts unique challenges (de Vaus 2008:255). The meaning of concepts and the operational definitions of variables may differ between nations or regions (Erikson 1966:xi), so the comparative researcher must consider how best to establish measurement equivalence (Markoff 2005:402). For example, the concept of being a *good son or daughter* refers to a much broader range of behaviors in China than in most Western countries (Ho 1996). Rates of physical disability cannot be compared between nations because standard definitions are lacking (Martin and Kinsella 1995:364–365). Individuals in different cultures may respond differently to the same questions (Martin and Kinsella 1995:385). Alternatively, different measures may have been used for the same concepts in different nations, and the equivalence of these measures may be unknown (van de Vijver and Leung 1997:9). The value of statistics for particular geographic units such as counties in the United States may vary over time simply because of changes in the boundaries of these units (Walters et al. 1997). Such possibilities should be considered, and any available opportunity should be taken to test for their effects.

Qualitative data can also be used as a primary tool for comparative research. The Human Relations Area Files (HRAF) Collection of Ethnography, also mentioned in Chapter 14, provides an extraordinary resource for qualitative comparative cross-sectional research (and, to a lesser extent, for qualitative comparative historical research) (Ember and Ember 2011). The HRAF was founded in 1949 as a corporation designed to facilitate cross-cultural research. The HRAF ethnography collection now contains more than 1 million pages of material from publications and other reports from about 400 different cultural, ethnic, religious, and national groups all over the world. The information is indexed by topic, in 710 categories and now made available electronically (if your school pays to maintain access to the HRAF). Figure 15.7 is an example of a page from an HRAF document that has been indexed for easy retrieval.

Most of the significant literature published on the chosen groups is included in the HRAF and used to prepare a standard summary about the group. Researchers can use these summaries and systematic searches for specific index terms to answer many questions about other social

FIGURE 15.7 ■ HRAF-Indexed Document

HRAF
eHRAF World Cultures
Home
Search
Browse
Help
Results Per Culture
Paragraph Results
Full Context
Full Context
First Page
Previous Page
Current Page
Next Page
Last Page
[introductory page 1]
197
198 (14 of 40)
199
224
< Previous Hit (Page 191)
198
Next Hit (Page 221) >
Region: Africa Culture: Amhara
Field Date: unknown Date Coverage: 1800s Place Coverage: Ethiopia
Print Selected Records
email Selected Records
Select
All
Though ultimately all authorized control which gives kibur flows from God, there are two major immediate sources or channels through which it is usually delegated to concrete individuals: through the emperor to the secular, military administrative hierarchy, and through the church to the clergy. The character of legitimate control is differentiated according to the channel through which it is legitimated. There are, thus, in traditional Amhara society, two partially independent and
Territorial hierarchy
Ethics
Priesthood
Document Table of Contents
Hoben, Allan, 1934- Social stratification in traditional Amhara society
Publication Information
Front Matter
Body
SOCIAL STRATIFICATION IN TRADITIONAL AMHARA SOCIETY
THE CONCEPT OF SOCIAL STRATIFICATION
THE SETTING

groups with the HRAF files, such as "What percentage of the world's societies practice polygyny?" and "Does punitive child training affect the frequency of warfare?" (Ember and Ember 2011).

Comparative Historical Research

The combination of historical analysis with comparisons between nations or other units often leads to the most interesting results. Historical social scientists may use comparisons between cases "to highlight the particular features of each case" (Skocpol 1984:370) or to identify general historical patterns across nations. A study of processes within one nation may therefore be misleading if important processes within the nation have been influenced by social processes that operate across national boundaries (Markoff 2005:403). For example, comparisons between nations may reveal that differences in political systems are much more important than voluntary decisions by individual political leaders (Rueschemeyer et al. 1992:31–36).

Comparative historical research can also help identify the causal processes at work within the nations or other entities (Lipset 1968:34; Skocpol 1984:374–386). Comparative historical research can result in historically conditional theory, in which the applicability of general theoretical propositions is linked to particular historical circumstances (Paige 1999). For example, James Mahoney (2001) explained the differential success of liberalizing reforms in Central American countries as the result of the particular ways in which liberal elites tried to implement the reforms. As summarized by Lange (2013:77–78), Guatemala and El Salvador tried to implement the reforms quickly by developing a militarized state, whereas Costa Rica used a gradual approach that could be implemented by a democratic regime. In Honduras and Nicaragua, internal pressure and external interventions stopped the liberal reforms entirely. These early events set political processes in these countries on different trajectories for many decades. In a similar way, Fallon, Swiss, and Viterna (2012) identified the influences shaping the extent of women's legislative representation in democratizing countries.

The great classical sociologists also used comparative methods, although their approach was less systematic. For example, Max Weber's (Bendix 1962:268) comparative sociology of religions contrasted Protestantism in the West, Confucianism and Taoism in China, Hinduism and Buddhism in India, and Ancient Judaism. As Bendix (1962) explained,

> His [Weber's] aim was to delineate religious orientations that contrasted sharply with those of the West, because only then could he specify the features that were peculiar to Occidental [Western] religiosity and hence called for an explanation . . . to bring out the distinctive features of each historical phenomenon. (p. 268)

So, for example, Weber concluded that the rise of Protestantism, with its individualistic approach to faith and salvation, was an important factor in the development of capitalism.

Mitchell Duneier's (2016) book, *Ghetto: The Invention of a Place, The History of an Idea,* uses this classical approach in historical comparative research. To understand the social origins of the concept of a ghetto and its influence on social behavior, Duneier compares the development of what came to be known as ghettos in medieval Venice, Nazi-occupied Warsaw, and then Chicago and New York's Harlem in different historical periods (Chicago in the 1940s and 1980s, and Harlem in the 1960s and 2000s). Although his primary focus is on ghettos in American cities, Duneier begins his analysis by comparing the history of 16th-century Venice's Jewish ghetto with the history of the Jewish ghetto in Nazi-occupied Warsaw. In both, he finds "the pernicious circular logic of the ghetto": "the consequences of ghettoization provided an apparent justification for the original condition" (p. 11).

But Duneier also identifies a striking contrast: Although daytime social circulation across the ghetto boundaries allowed a wide range of social institutions to continue in spite of religious prejudices, the barbed wire barriers around Warsaw's ghetto facilitated a complete segregation that was designed to hasten its occupants' deaths. Segregation, Duneier concludes, can be compatible "with wide variations in both control and flourishing" (p. 220). When he turns to Chicago and New York, relying primarily on the work of local Black scholars, Duneier finds in earlier years evidence of more functional communities that are then decimated by growing opportunities for middle-class Blacks to move to the suburbs and by increasingly punitive criminal justice policies. Duneier concludes that "the ghetto is an expression of societal power" reflecting prejudice against Blacks as well as "ongoing external domination and neglect" that spans generations (pp. 222–225).

Much modern comparative historical research takes a more systematic approach to comparisons, usually focusing on a detailed sequence of events that might have influenced an outcome in the present. Some studies collect quantitative longitudinal data about a number of nations and then use these data to test hypotheses about influences on national characteristics. (Theda Skocpol [1984:375] terms this *analytic historical sociology.*) Others compare the histories or particular historical experiences of nations in a narrative form, noting similarities and differences and inferring explanations for key national events (*interpretive historical sociology* in Skocpol's terminology [1984:368]). Either quantitative or qualitative data can be used.

There are several stages for a systematic, qualitative, comparative historical study (Ragin 1987:44–52; Rueschemeyer et al. 1992:36–39):

1. Specify a theoretical framework and identify key concepts or events that should be examined to explain a phenomenon.
2. Select cases (such as nations) that vary in terms of the key concepts or events.
3. Identify similarities and differences between the cases in these key concepts or events and the outcome to be explained.
4. Propose a causal explanation for the historical outcome and check it against the features of each case. The criterion of success in this method is to explain the outcome for each case, without allowing deviations from the proposed causal pattern.

Dietrich Rueschemeyer et al. (1992) used a comparative historical method to explain why some nations in Latin America (excluding Central America) developed democratic politics, whereas others became authoritarian or bureaucratic–authoritarian states. First, Rueschemeyer et al. developed a theoretical framework that gave key attention to the power of social classes, state (government) power, and the interaction between social classes and the government. The researchers then classified the political regimes in each nation over time (see Table 15.3). Next, they noted how each nation varied over time relative to the variables they had identified as potentially important for successful democratization.

TABLE 15.3 ■ Classification of Regimes Over Time

	Constitutional Oligarchic	Authoritarian; Traditional, Populist, Military, or Corporatist	Restricted Democrat	Fully Democratic	Bureaucratic–Authoritarian
Argentina	before 1912	1930–46	1958–62	1912–30	1966–73
		1951–55	1963–66	1946–51	1976–83
		1955–58		1973–76	
		1962–63		1983–90	
Bolivia	before 1930	1930–52	1982–90	1952–64	
		1964–82			
Brazil	before 1930	1930–45	1945–64		1964–85
			1985–90		
Chile	before 1920	1924–32	1920–24	1970–73	1973–89
			1932–70		
			1990		
Colombia	before 1936	1949–58	1936–49		
			1958–90		
Ecuador	1916–25	before 1916	1948–61		
		1925–48	1978–90		
		1961–78			
Mexico		up to 1990			
Paraguay		up to 1990			
Peru		before 1930	1939–48	1980–90	
		1930–39	1956–62		
		1948–56	1963–68		

	Constitutional Oligarchic	Authoritarian; Traditional, Populist, Military, or Corporatist	Restricted Democrat	Fully Democratic	Bureaucratic–Authoritarian
		1962–63			
		1968–80			
Uruguay		before 1903	1903–19	1919–33	1973–84
		1933–42		1942–73	
				1984–90	
Venezuela		before 1935–45	1958–68	1945–48	
		1948–58		1968–90	

Source: Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens. 1992. *Capitalist Development and Democracy.* Reprinted with permission from the University of Chicago Press.

Their analysis identified several conditions for initial democratization: consolidation of state power (ending overt challenges to state authority), expansion of the export economy (reducing conflicts over resources), industrialization (increasing the size and interaction of middle and working classes), and some agent of political articulation of the subordinate classes (which could be the state, political parties, or mass movements). Historical variation in these conditions was then examined in detail.

When geographic units such as nations are sampled for comparative purposes, it is assumed that the nations are independent of each other in the variables examined. Each nation can then be treated as a separate case for identifying possible chains of causes and effects. However, in a very interdependent world, this assumption may be misplaced—nations may develop as they do because of how other nations are developing (and the same can be said of cities and other units). As a result, comparing the particular histories of different nations may overlook the influence of global culture, international organizations, or economic dependency—just the type of influence identified in Frank et al.'s study of environmental protectionism that was the focus of Figure 15.4 (Skocpol 1984:384; compare Chase-Dunn and Hall 1993). These common international influences may cause the same pattern of changes to emerge in different nations; looking within the history of these nations for the explanatory influences would lead to spurious conclusions (de Vaus 2008:258). The possibility of such complex interrelations should always be considered when evaluating the plausibility of a causal argument based on a comparison between two apparently independent cases (Jervis 1996).

RESEARCH IN THE NEWS
THE SECRET HISTORY OF WOMEN IN CODING

In Britain, women operated some of the first code-breaking computational machines during World War II. In the United States, by 1960, more than one quarter of computer programmers were women. "We were a bunch of geeks," one recalled (Thompson 2019:40). In the 1970s, about equal numbers of men and women reported an interest in a career in computer coding, and by the 1983–1984 academic year, women represented 37.1% of all students graduating with degrees in computer and information science. But by 2010, that percentage had

dropped to 17.6. What happened? Research by social scientist Jane Margolis at the University of California Los Angeles provided an answer. Until the advent of personal computers in the 1980s, all students were exposed to computers for the first time on campus; the machines at the time were big and expensive. However, with the development and proliferation of personal computers, teenagers began to learn how to use computers at home and increasingly came to college with advanced skills. And parents who bought personal computers for the family mostly did them for their sons, who were then helped by their fathers. Girls got the message that they should stick with dolls and toy kitchens. By the time they entered college, male students saw themselves as participating in a masculine culture based on technical skills and were given much more attention by their professors than those female students still bold enough to sign up for programming classes.

For Further Thought?

1. How have you viewed the gender imbalance in computers? Has it seemed "natural" to you? What additional historical evidence would you propose collecting to test the portion of Margolis's explanation for that imbalance that is described in this vignette?
2. Suggest a qualitative or quantitative comparative historical research project that would help to explain gender imbalance in computer science on a global scale.

News source: Thompson, Clive. 2019. "The Secret History of Women in Coding." *The New York Times Magazine*, February 17:38–45, 54–55.

Comparative Case Study Designs

Some comparative researchers use a systematic method for identifying causes that owes its origins to the English philosopher John Stuart Mill (1872). One approach that Mill developed was called the **method of agreement**. The core of this approach is the comparison of nations (*cases*) for similarities and differences on potential causal variables and the phenomenon to be explained. As comparative historian, Skocpol (1979:36) explains, researchers who use this method should "try to establish that several cases having in common the phenomenon one is trying to explain also have in common a set of causal factors, although they vary in other ways that might have seemed causally relevant."

For example, suppose three countries that have all developed democratic political systems are compared through four socioeconomic variables hypothesized by different theories to influence democratization (see Table 15.4). If the countries differ in three of the variables but are similar in the fourth, this is evidence that the fourth variable influences democratization. In Table 15.4,

TABLE 15.4 ■ John Stuart Mill's Method of Agreement (Hypothetical Cases and Variables)

Variable	Case 1	Case 2	Case 3
Importance of peasant agriculture	Different	Different	Different
Expanding industrial base	Different	Same	Same
Rising educational levels	Different	Different	Different
Expanding middle class	Same	Same	Same
Democratization (outcome)	Same	Same	Same

Source: Adapted from Skocpol (1984:379).

the method of agreement would lead the analyst to conclude that an expanding middle class was a cause of the democratization experienced in all three countries. The focus of the method of agreement is actually on identifying a similarity between cases that differ in many respects, so this approach is also called the *most different case studies* method.

The second approach Mill developed was the **method of difference**. Again, in the words of comparative historian Skocpol (1979),

> One can contrast the cases in which the phenomenon to be explained and the hypothesized causes are present to other cases in which the phenomenon and the causes are both absent, but which are otherwise as similar as possible to the positive cases. (p. 36)

The method of difference approach is represented in Table 15.5. In this example, "moderate income disparities" are taken to be the cause of democratization because the country that didn't democratize differs in this respect from the country that did democratize. These two countries are similar with respect to other potential influences on democratization. The argument could be improved by adding more positive and negative cases. The focus of the method of difference is on identifying a difference between cases that are similar in other respects, so this approach is also called the *most similar case studies* method.

TABLE 15.5 ■ John Stuart Mill's Method of Difference (Hypothetical Cases and Variables)

Country A (Positive Case)	Country B (Negative Case)
Economic development	Economic development
Two-party system begun	Two-party system begun
Proportional representation	Proportional representation
Moderate income disparities	Extreme income disparities
Democratization	No democratization

Source: Adapted from Skocpol and Somers (1979:80).

The method of agreement and method of difference approaches can also be combined, "by using at once several positive cases along with suitable negative cases as contrasts" (Skocpol 1979:37). This is the approach that Skocpol (1979) used in her classic book about the French, Russian, and Chinese revolutions, *States and Social Revolutions.* Table 15.6 summarizes part of her argument about the conditions for peasant insurrections, based on a careful historical review. In this figure, Skocpol (1979:156) shows how the three countries that experienced revolutions (France, Russia, and China) tended to have more independent peasants and more autonomy in local politics than did three contrasting countries (Prussia/Germany, Japan, and England) that did not experience social revolutions.

Cautions for Comparative Methods

Of course, ambitious methods that compare different countries face many complications. The features of the cases selected for comparison have a large impact on the researcher's ability to identify influences. Cases should be chosen for their difference on key factors hypothesized to influence the outcome of interest and their similarity on other, possibly confounding, factors

TABLE 15.6 ■ Methods of Agreement and Difference Combined: Conditions for Peasant Insurrections

Country	Agrarian Class Structures	Local Politics
France	Peasant smallholders own 30% to 40% of land; work 80% in small plots. Individual property established, but peasant community opposes seigneurs, who collect dues.	Villages relatively autonomous under supervision of royal officials.
Russia	Peasants own 60%+ and rent more; control process of production on small plots; pay rents and redemption payments. Strong community based on collective ownership.	Village sovereign under control of tzarist bureaucracy.
China	Peasants own 50% and work virtually all land in small plots. Pay rents to gentry. No peasant community.	Gentry landlords, usurers, and literati dominate local organizational life; cooperate with Imperial officials.
Contrasts		
Prussia/Germany	West of Elbe: resembles France. East of Elbe: large estates worked by laborers and peasants with tiny holdings and no strong communities.	Junker landlords are local agents of bureaucratic state; dominate local administration and policing.
Japan	Communities dominated by rich peasants.	Strong bureaucratic controls over local communities.
England	Landed class owns 70%. Peasantry polarizing between yeomen farmers and agricultural laborers. No strong peasant community.	Landlords are local agents of monarchy; dominate administration and policing.

Source: Skocpol (1979).

(Skocpol 1984:383). For example, to understand how industrialization influences democracy, you would need to select cases for comparison that differ in industrialization so that you could then see if they differ in democratization (King, Keohane, and Verba 1994:148–152). Nonetheless, relying on just a small number of cases for comparisons introduces uncertainty into the conclusions (de Vaus 2008:256).

And what determines whether cases are similar and different in certain respects? In many comparative analyses, the values of continuous variables are dichotomized. For example, nations may be coded as *democratic* or *not democratic* or as having *experienced revolution* or *not experienced revolution*. The methods of agreement and difference that I have just introduced presume these types of binary (dichotomous) distinctions. However, variation in the social world often involves degrees of difference rather than all or none distinctions (de Vaus 2008:255). Some countries may be partially democratic and some countries may have experienced a limited revolution. At the individual level, you know that distinctions such as *rich* and *poor* or *religious* and *not religious* reflect differences on underlying continua of wealth and religiosity. So the use of dichotomous distinctions in comparative analyses introduces an imprecise and somewhat arbitrary element into the analysis (Lieberson 1991). For some comparisons, however, qualitative distinctions, such

as *simple majority rule* or *unanimity required,* may capture the important differences between cases better than quantitative distinctions. We don't want to simply ignore important categorical considerations such as this in favor of *degree of majority rule* or some other underlying variable (King et al. 1994:158–163). Careful discussion of the bases for making distinctions is an important first step in any comparative historical research (also see Ragin 2000).

The focus on comparisons between nations may itself be a mistake for some analyses. National boundaries often do not correspond to key cultural differences, so comparing subregions within countries or larger cultural units that span multiple countries may make more sense for some analyses (de Vaus 2008:258). Comparing countries that have fractured along cultural or religious divides simply by average characteristics would obscure many important social phenomena.

With cautions such as these in mind, the combination of historical and comparative methods allows for rich descriptions of social and political processes in different nations or regions as well as for causal inferences that reflect a systematic, defensible weighing of the evidence. Data of increasingly good quality are available for a rapidly expanding number of nations, creating many opportunities for comparative research. We cannot expect one study comparing the histories of a few nations to control adequately for every plausible alternative causal influence, but repeated investigations can refine our understanding and lead to increasingly accurate causal conclusions (King et al. 1994:33).

DEMOGRAPHIC ANALYSIS

The social processes that are the focus of historical and comparative research are often reflected in and influenced by changes in the makeup of the population being studied. For example, the plummeting birth rates in European countries can influence the politics of immigration in those countries, their living standards, the character of neighborhoods, and national productivity (Bruni 2002). **Demography** is the field that studies these dynamics. Demography is the statistical and mathematical study of the size, composition, and spatial distribution of human populations and how these features change over time. Demographers explain population change through five processes: (1) fertility, (2) mortality, (3) marriage, (4) migration, and (5) social mobility (Bogue 1969:1).

Demographers obtain data from a census of the population (see Chapter 5) and from registries—records of events, such as births, deaths, migrations, marriages, divorces, diseases, and employment (Anderton, Barrett, and Bogue 1997:54–79; Baum 1993)—then compute various statistics from these data to facilitate description and analysis (Wunsch and Termote 1978). To use these data, you need to understand how they are calculated and the questions they answer. Four concepts are key to understanding and using demographic methods: population change, standardization of population numbers, the demographic bookkeeping equation, and population composition.

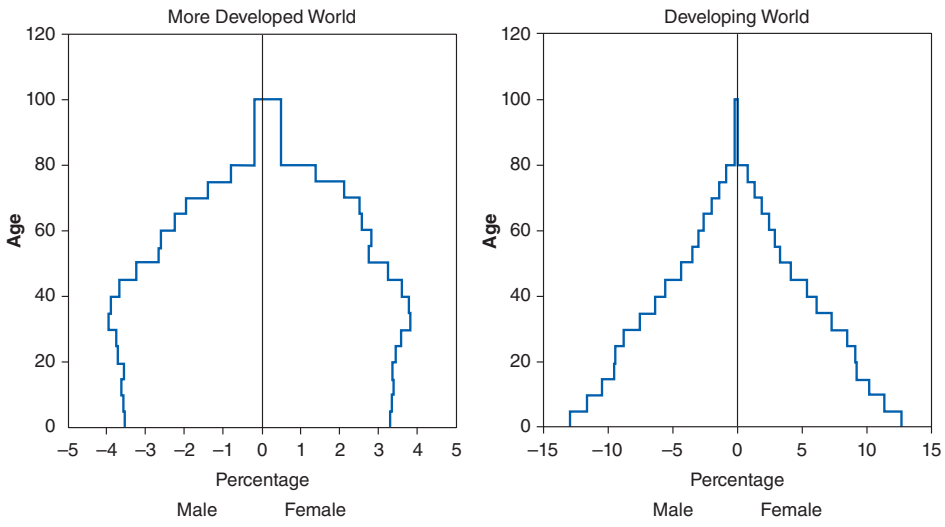
Population change is a central concept in demography. The absolute population change is calculated simply as the difference between the population size in one census minus the population size in an earlier census. This measure of absolute change is of little value, however, because it does not consider the total size of the population that was changing (Bogue 1969:32–43). A better measure is the *intercensal percent change*, which is the absolute change in population between two censuses divided by the population size in the earlier census (and multiplied by 100 to obtain a percentage). With the percent change statistic, we can meaningfully compare the growth in two or more nations that differ markedly in size (as long as the intercensal interval does not vary between the nations) (White 1993:1–2).

Standardization of population numbers is a key concern of demographic methods (Gill, Glazer, and Thernstrom 1992:478–482; Rele 1993). The calculation of intercensal percent change, as described, is a good example. To make meaningful comparisons between nations and over time, numbers that describe most demographic events must be adjusted for the size of the population at risk for the event. For example, the fertility rate is calculated as the ratio of the number of births to women of childbearing age to the total number of women in this age range (multiplied by 1,000). Unless we make such adjustments, we will not know if a nation with a much higher number of births or deaths in relation to its total population size simply has more women in the appropriate age range or has more births per "eligible" woman.

The *demographic bookkeeping* (or *balancing*) *equation* is used to identify the four components of population growth during a time interval (P_2P_1): births (B), deaths (D), and in-migration (M_i) and out-migration (M_o). The equation is written as follows: $P_2 = P_1 + (B - D) + (M_i - M_o)$. That is, population at a given point in time is equal to the population at an earlier time plus the excess of births over deaths during the interval and the excess of in-migration over out-migration (White 1993:1–4). Whenever you see population size or change statistics used in a comparative analysis, you will want to ask yourself whether it is also important to know which component in the equation was responsible for the change over time or for the difference between countries (White 1993:1–4).

Population composition refers to a description of a population in terms of basic characteristics such as age, race, sex, or marital status (White 1993:1–7). Descriptions of population composition at different times or in different nations can be essential for understanding social dynamics identified in historical and comparative research. For example, Figure 15.8 compares the composition of the population in more developed and developing regions of the world by age and sex in 1995, using United Nations data. By comparing these *population pyramids*, we see that children constitute a much greater proportion of the population in less developed regions. The more developed regions' population pyramid also shows the greater proportion of women at older ages and the post–World War II baby boom bulge in the population.

FIGURE 15.8 ■ Population Pyramids for More Developed and Developing Regions of the World: 1995*



Source: Bogue, Donald J., Eduardo E. Arriaga, and Douglas L. Anderton. 1993. *Readings in Population Research Methodology, vol. 1, Basic Tools.* Chicago: Social Development Center, for the United Nations Populations Fund.

Demographic analysis can be an important component of historical research (Bean, Mineau, and Anderton 1990), but problems of data quality must be evaluated carefully (Vaessen 1993). The hard work that can be required to develop demographic data from evidence that is hundreds of years old does not always result in worthwhile information. The numbers of people for which data are available in particular areas may be too small for statistical analysis, data that are easily available (e.g., a list of villages in an area) may not provide the information that is important (e.g., population size), and lack of information on the original data collection procedures may prevent assessment of data quality (Hollingsworth 1972:77).

CONTENT ANALYSIS

How are medical doctors regarded in U.S. culture? Do newspapers use the term *schizophrenia* in a way that reflects what this serious mental illness actually involves? Does the portrayal of men and women in video games reinforce gender stereotypes? Are the body images of male and female college students related to their experiences with romantic love? If you are concerned with understanding culture, attitudes toward mental illness, or gender roles, you'll probably find these to be important research questions. You now know that you could probably find data about each of these issues for a secondary data analysis, but in this section, I would like to introduce procedures for analyzing a different type of data that awaits the enterprising social researcher. **Content analysis** is "the systematic, objective, quantitative analysis of message characteristics" and is a method particularly well suited to the study of popular culture and many other issues concerning human communication (Neuendorf 2002:1).

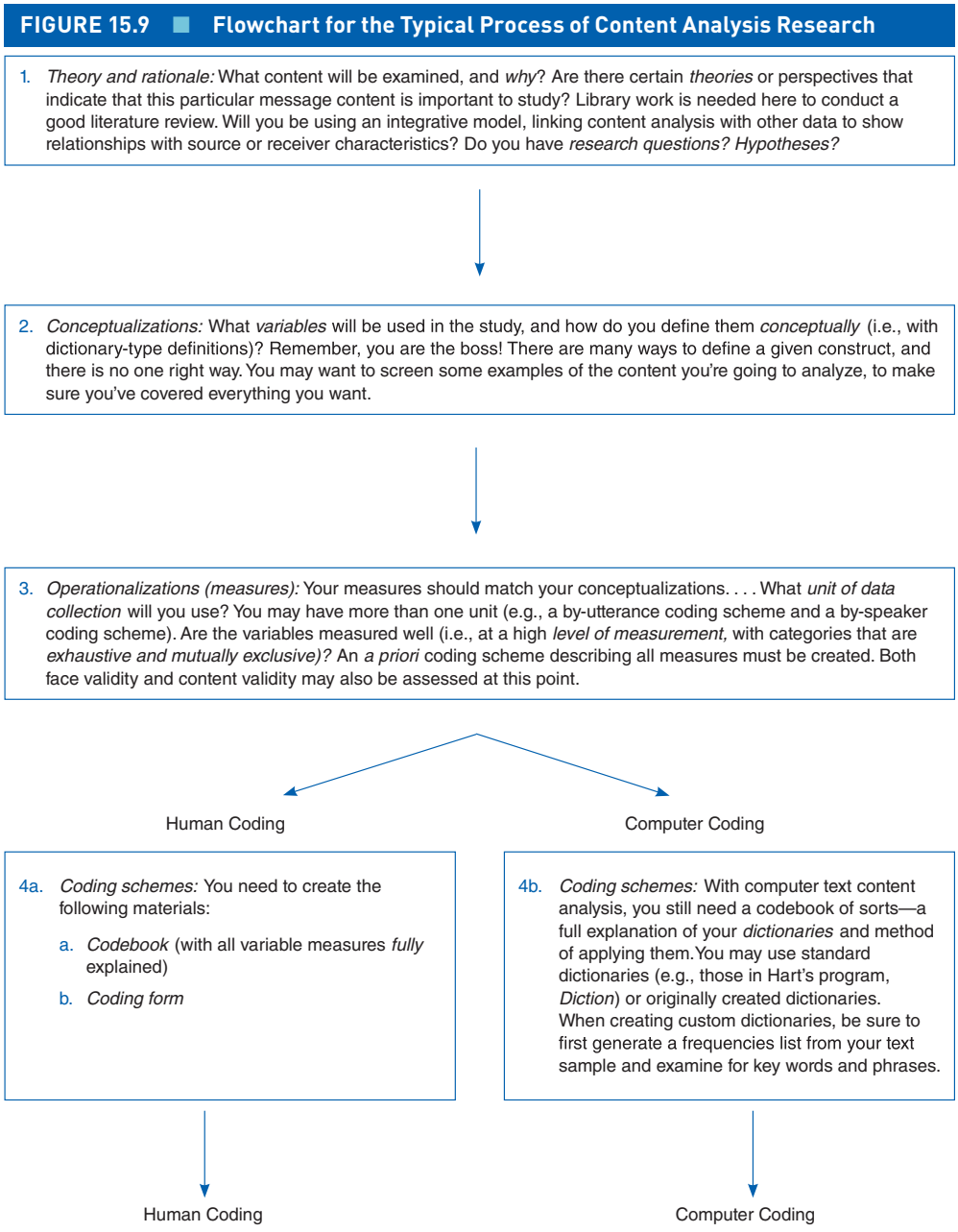
Content analysis can be used in historical and comparative research but is also useful in studies of communication with any type of media. Therefore, like most historical and comparative methods, content analysis can be called an *unobtrusive method* that does not need to involve interacting with live people. Content analysis methods usually begin with text, speech broadcasts, or visual images. The content analyst develops procedures for coding various aspects of the textual, aural (spoken), or visual material and then analyzes this coded content.

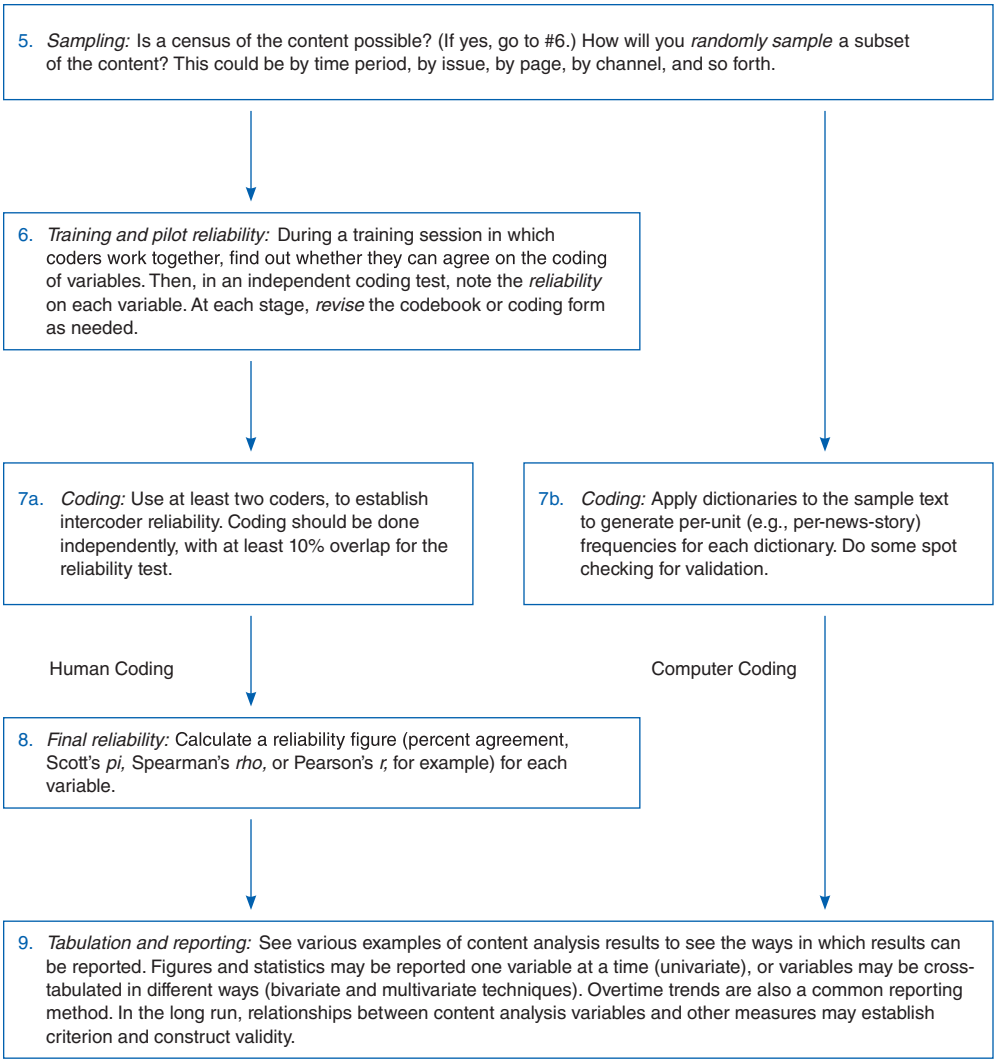
The goal of content analysis is to develop inferences from human communication in any of its forms, including books, articles, magazines, songs, films, and speeches (Weber 1990:9). You can think of content analysis as a "survey" of some documents or other records of communication—a survey with fixed-choice responses that produce quantitative data. This method was first applied to the study of newspaper and film content and then developed systematically for the analysis of Nazi propaganda broadcasts in World War II. Since then, content analysis has been used to study historical documents, records of speeches, and other "voices from the past" as well as media of all sorts (Neuendorf 2002:31–37). The same techniques can now be used to analyze blog sites, wikis, and other text posted on the Internet (Gaiser and Schreiner 2009:81–90). Content analysis techniques are also used to analyze responses to open-ended survey questions.

Content analysis bears some similarities to qualitative data analysis because it involves coding and categorizing text and identifying relationships between constructs identified in the text. However, because it usually is conceived as a quantitative procedure, content analysis overlaps with qualitative data analysis only at the margins—the points where qualitative analysis takes on quantitative features or where content analysis focuses on qualitative features of the text. This distinction becomes fuzzy, however, because content analysis techniques can be used with all forms of messages, including visual images, sounds, and interaction patterns, as well as written text (Neuendorf 2002:24–25).

The various steps in a content analysis are represented in the flowchart in Figure 15.9. Note that the steps are comparable to the procedures in quantitative survey research. Use this flowchart as a checklist when you design or critique a content analysis project.

FIGURE 15.9 ■ Flowchart for the Typical Process of Content Analysis Research


1. Theory and rationale: What content will be examined, and why? Are there certain theories or perspectives that indicate that this particular message content is important to study? Library work is needed here to conduct a good literature review. Will you be using an integrative model, linking content analysis with other data to show relationships with source or receiver characteristics? Do you have research questions? Hypotheses?
2. Conceptualizations: What variables will be used in the study, and how do you define them conceptually (i.e., with dictionary-type definitions)? Remember, you are the boss! There are many ways to define a given construct, and there is no one right way. You may want to screen some examples of the content you're going to analyze, to make sure you've covered everything you want.
3. Operationalizations (measures): Your measures should match your conceptualizations. . . . What unit of data collection will you use? You may have more than one unit (e.g., a by-utterance coding scheme and a by-speaker coding scheme). Are the variables measured well (i.e., at a high level of measurement, with categories that are exhaustive and mutually exclusive)? An a priori coding scheme describing all measures must be created. Both face validity and content validity may also be assessed at this point.
Human Coding
Computer Coding
4a. Coding schemes: You need to create the following materials:
a. Codebook (with all variable measures fully explained)
b. Coding form
4b. Coding schemes: With computer text content analysis, you still need a codebook of sorts—a full explanation of your dictionaries and method of applying them. You may use standard dictionaries (e.g., those in Hart's program, Diction) or originally created dictionaries. When creating custom dictionaries, be sure to first generate a frequencies list from your text sample and examine for key words and phrases.
Human Coding
Computer Coding

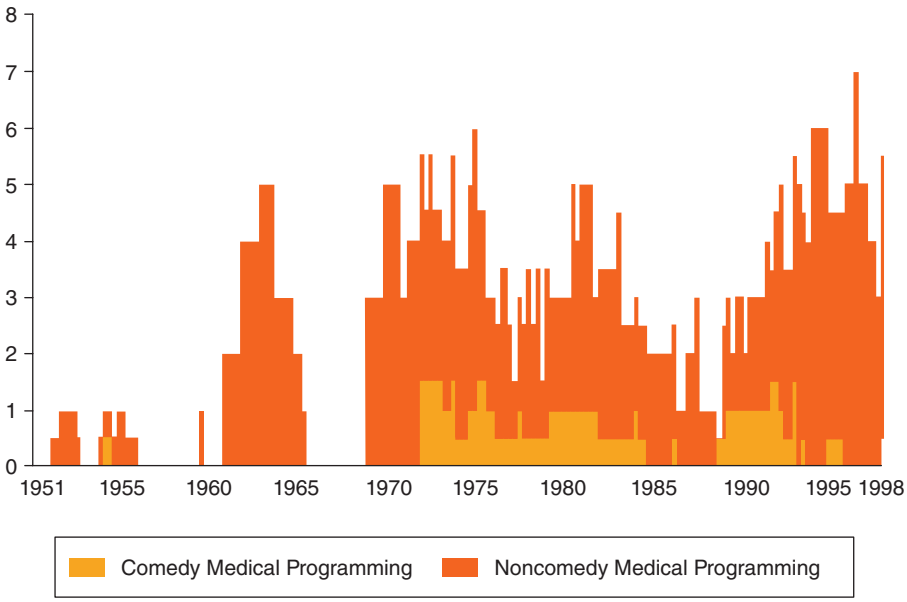

5. Sampling: Is a census of the content possible? (If yes, go to #6.) How will you randomly sample a subset of the content? This could be by time period, by issue, by page, by channel, and so forth.
6. Training and pilot reliability: During a training session in which coders work together, find out whether they can agree on the coding of variables. Then, in an independent coding test, note the reliability on each variable. At each stage, revise the codebook or coding form as needed.
7a. Coding: Use at least two coders, to establish intercoder reliability. Coding should be done independently, with at least 10% overlap for the reliability test.
7b. Coding: Apply dictionaries to the sample text to generate per-unit (e.g., per-news-story) frequencies for each dictionary. Do some spot checking for validation.
Human Coding
Computer Coding
8. Final reliability: Calculate a reliability figure (percent agreement, Scott's pi, Spearman's rho, or Pearson's r, for example) for each variable.
9. Tabulation and reporting: See various examples of content analysis results to see the ways in which results can be reported. Figures and statistics may be reported one variable at a time (univariate), or variables may be cross-tabulated in different ways (bivariate and multivariate techniques). Overtime trends are also a common reporting method. In the long run, relationships between content analysis variables and other measures may establish criterion and construct validity.

Source: Neuendorf (2002).

Kimberly Neuendorf's (2002:3) analysis of medical prime-time network television programming introduces the potential of content analysis. As Figure 15.10 shows, medical programming has been dominated by noncomedy shows, but there have been two significant periods of comedy medical shows—during the 1970s and early 1980s and then again in the early 1990s. It took a qualitative analysis of medical show content to reveal that the 1960s shows represented a very distinct "physician-as-God" era, which shifted to a more human view of the medical profession in the 1970s and 1980s. This era has been followed, in turn, by a mixed period that has had no dominant theme.

Content analysis is useful for investigating many questions about the social world. To illustrate its diverse range of applications, I will use in the next sections Neuendorf's (2002) analysis of TV programming; Matthias A. Gerth and Gabriele Siegert's (2012) analysis of news coverage of an immigrant naturalization campaign; Kenneth Duckworth's, John Halpern's, Chris Gillespie's, and my (2003) analysis of newspaper articles; Karen Dill and Kathryn Thill's (2007) analysis of video game characters; and Suman Ambwani and Jaine Strauss's (2007) analysis of student responses to open-ended survey questions. These examples will demonstrate that the

FIGURE 15.10 ■ Medical Prime-Time Network Television Programming, 1951–1998



Source: Neuendorf, Kimberly A. 2002. *The Content Analysis Guidebook*. Thousand Oaks, CA: SAGE.

units that are "surveyed" in a content analysis can range from newspapers, books, or TV shows to persons referred to in other communications, themes expressed in documents, or propositions made in different statements.

Identify a Population of Documents or Other Textual Sources

The population of documents or other textual sources should be selected so that it is appropriate to the research question of interest. Perhaps the population will be all newspapers published in the United States, college student newspapers, nomination speeches at political party conventions, or "state of the nation" speeches by national leaders. Books or films are also common sources for content analysis projects. Often, a comprehensive archive can provide the primary data for the analysis (Neuendorf 2002:76–77). For a fee, the LexisNexis service makes a large archive of newspapers available for analysis. For her analysis of prime-time programming since 1951, Neuendorf (2002:3–4) used a published catalog of all TV shows. For my analysis with Duckworth and others (2003:1402) of newspapers' use of the terms *schizophrenia* and *cancer*, I requested a sample of articles from the LexisNexis national newspaper archive. Gerth and Siegert (2012) focused on TV and newspaper stories during a 14-week Swiss political campaign, and Dill and Thill (2007:855–856) turned to video game magazines for their analysis of the depiction of gender roles in video games. For their analysis of gender differences in body image and romantic love, Ambwani and Strauss (2007:15) surveyed students at a small midwestern liberal arts college.

Determine the Units of Analysis

The units of analysis could be items such as newspaper articles, whole newspapers, speeches, or political conventions, or they could be more microscopic units, such as words, interactions, time periods, or other bits of a communication (Neuendorf 2002:71). The content analyst has

to decide what units are most appropriate to the research question and how the communication content can be broken into those units. If the units are individual issues of a newspaper, in a study of changes in news emphases, this step may be relatively easy. However, if the units are most appropriately the instances of interaction between characters in a novel or a movie, in a study of conflict patterns between different types of characters, it will require a careful process of testing to determine how to define operationally the specific units of interaction (Weber 1990:39–40).

Units of analysis varied across the five content analysis projects I have introduced. The units of analysis for Neuendorf (2002:2) were "the individual medically oriented TV program"; for Duckworth et al. (2003:1403), they were newspaper articles; for Gerth and Siegert (2012:288), they were arguments made in media stories; and for Dill and Thill (2007:856), they were images appearing in magazine articles. The units of analysis for Ambwani and Strauss (2007:15) were individual students.

Select a Sample of Units From the Population

The simplest strategy might be to select a simple random sample of documents. However, a stratified sample might be needed to ensure adequate representation of community newspapers in large and small cities or of weekday and Sunday papers or of political speeches during election years and in off years (seeChapter 4) (Weber 1990:40–43). Nonrandom sampling methods have also been used in content analyses when the entire population of interest could not be determined (Neuendorf 2002:87–88).

The selected samples in our five content analysis projects were diverse. In fact, Neuendorf (2002:2) included the entire population of medically oriented TV programs between 1951 and 1998. For my content analysis with Duckworth (Duckworth et al. 2003), I had my student, Chris Gillespie, draw a stratified random sample of 1,802 articles published in the five U.S. newspapers with the highest daily circulation in 1996 to 1997 in each of the four regions identified in the LexisNexis database, as well as the two high-circulation national papers in the database, *The New York Times* and *USA Today* (pp. 1402–1403).

Because individual articles cannot be sampled directly in the LexisNexis database, a random sample of days was drawn first. All articles using the terms *schizophrenia* or *cancer* (or several variants of these terms) were then selected from the chosen newspapers on these days. Gerth and Siegert (2012:285) selected 24 different newspapers and five TV news programs that targeted the population for the campaign, and they then coded 3,570 arguments made in them about the campaign during its 14 weeks. Dill and Thill (2007:855–856) used all images in the current issues (as of January 2006) of the six most popular video game magazines sold on Amazon.com. Ambwani and Strauss (2007:15) used an availability sampling strategy, with 220 students from Introductory Psychology and a variety of other sources.

Design Coding Procedures for the Variables to Be Measured

Designing coding procedures requires deciding what variables to measure using the unit of text to be coded, such as words, sentences, themes, or paragraphs. Then, the categories into which the text units are to be coded must be defined. These categories may be broad, such as *supports democracy*, or more narrow, such as *supports universal suffrage.* Reading or otherwise reviewing some of the documents or other units to be coded is an essential step in thinking about variables that should be coded and in developing coding procedures. Development of clear instructions and careful training of coders are essential.

As an example, Figure 15.11 is a segment of the coding form that I developed for a content analysis of union literature that I collected during a mixed-methods study of union political

processes (Schutt 1986). My sample was of 362 documents: all union newspapers and a stratified sample of union leaflets given to members during the years of my investigation. My coding scheme included measures of the source and target for the communication, as well as measures of concepts that my theoretical framework indicated were important in organizational development: types of goals, tactics for achieving goals, organizational structure, and forms of participation. The analysis documented a decline in concern with client issues and an increase in focus on organizational structure, which were trends that also emerged in interviews with union members.

Developing reliable and valid coding procedures deserves special attention in a content analysis, for it is not an easy task. The meaning of words and phrases is often ambiguous. Homographs create special problems (words such as *mine* that have different meanings in different contexts), as do many phrases that have special meanings (such as *point of no return*) (Weber 1990:29–30). As a result, coding procedures cannot simply categorize and count words; text segments in which the words are embedded must also be inspected before codes are finalized. Because different coders may perceive different meanings in the same text segments, explicit coding rules are required to ensure coding consistency. Special dictionaries can be developed to keep track of how the categories of interest are defined in the study (Weber 1990:23–29).

FIGURE 15.11 ■ Union Literature Coding Form

I. *Preliminary Codes*

1. Document # _____
2. Date _____

   mo yr
3. Length of text _____ pp. (round up to next 1/4 page; count legal size as 1.25)
4. Literature Type
   1. General leaflet for members/employees
   2. Newspaper/Newsletter article
   3. Rep Council motions
   4. Other material for Reps, Stewards, Delegates (e.g., budget, agenda)
   5. Activity reports of officers, President's Report
   6. Technical information-filing grievances, processing forms
   7. Buying plans/Travel packages
   8. Survey Forms, Limited Circulation material (correspondence)
   9. Non-Union
   10. Other _____ (specify)

4A. If newspaper article

Position

1. Headline story
2. Other front page
3. Editorial
4. Other

4B. If Rep Council motion

Sponsor

1. Union leadership
2. Office
3. Leadership faction
4. Opposition faction
5. Other

5. Literature content—Special issues
    1. First strike (1966)
    2. Second strike (1967)
    3. Collective bargaining (1977)
    4. Collective bargaining (1979)
    5. Election/campaign literature
    6. Affiliation with AFSCME/SEIU/other national union
    7. Other

II. *Source and Target*

6. Primary source (code in terms of those who prepared this literature for distribution).
    1. Union-newspaper (Common Sense; IUPAE News)
    2. Union-newsletter (Info and IUPAE Bulletin)
    3. Union-unsigned
    4. Union officers
    5. Union committee
    6. Union faction (the Caucus; Rank-and-Filers; Contract Action, other election slate; PLP News; Black Facts)
    7. Union members in a specific work location/office
    8. Union members—other
    9. Dept. of Public Aid/Personnel
    10. DVR/DORS
    11. Credit Union
    12. Am. Buyers' Assoc.
    13. Other nonunion

7. Secondary source (use for lit. at least in part reprinted from another source, for distribution to members)
    1. Newspaper—general circulation
    2. Literature of other unions, organizations
    3. Correspondence of union leaders
    4. Correspondence from DPA/DVR-DORS/Personnel
    5. Correspondence from national union
    6. Press release
    7. Credit Union, Am. Buyers'
    8. Other _____ (specify)
    9. None

8. Primary target (the audience for which the literature is distributed)
    1. Employees—general (if mass-produced and unless otherwise stated)
    2. Employees-DVR/DORS
    3. Union members (if refers only to members or if about union elections)
    4. Union stewards, reps, delegates committee
    5. Non-unionized employees (recruitment lit, etc.)
    6. Other _____ (specify)
    7. Unclear

(*Continued*)

FIGURE 15.11 ■ Union Literature Coding Form (*Continued*)

III. Issues

A. Goal

B. Employee conditions/benefits (Circle up to 5)

1. Criteria for hiring
2. Promotion
3. Work out of Classification, Upgrading
4. Step increases
5. Cost-of-living, pay raise, overtime pay, "money"
6. Layoffs (nondisciplinary); position cuts
7. Workloads, Redeterminations, "30 for 40," GA Review
8. Office physical conditions, safety
9. Performance evaluations
10. Length of workday
11. Sick benefits/leave—holidays, insurance, illness, vacation, voting time
12. Educational leave
13. Grievances—change in procedures
14. Discrimination (race, sex, age, religion, national origin)
15. Discipline—political (union-related)
16. Discipline—performance, other
17. Procedures with clients, at work
18. Quality of work, "worthwhile jobs"—other than relations with clients

Source: Reprinted by permission from Organization in a Changing Environment: Unionization of Welfare Employees, edited by Russell K. Schutt, the State University of New York Press © 1986, State University of New York. All rights reserved.

After coding procedures are developed, their reliability should be assessed by comparing different coders' codes for the same variables. Computer programs for content analysis can enhance reliability by facilitating the consistent application of text-coding rules (Weber 1990:24–28). Validity can be assessed with a construct validation approach by determining the extent to which theoretically predicted relationships occur (see Chapter 4).

Neuendorf's (2002:2) analysis of medical programming measured two variables that did not need explicit coding rules: length of show in minutes and the year(s) the program was aired. She also coded shows as comedies or noncomedies, as well as medical or not, but she does not report the coding rules for these distinctions. We provided a detailed description of coding procedures in our analysis of newspaper articles that used the terms *schizophrenia* or *cancer* (Duckworth et al. 2003). This description also mentions our use of a computerized text-analysis program and procedures for establishing measurement reliability:

> Content coding was based on each sentence in which the key term was used. Review of the full text of an article resolved any potential ambiguities in proper assignment of codes. Key terms were coded into one of eight categories: metaphor, obituary, first person or human interest, medical news, prevention or education, incidental, medically inappropriate, and charitable foundation. Fifty-seven of the 913 articles that mentioned schizophrenia but none of those that mentioned cancer were too ambiguous to be

> coded into one of these defined categories and so were excluded from final comparisons. Coding was performed by a trained graduate assistant using QSR's NUD*IST program. In questionable cases, final decisions were made by consensus of the two psychiatrist coauthors. A random subsample of 100 articles was also coded by two psychiatry residents, who, although blinded to our findings, assigned the same primary codes to 95 percent of the articles. (p. 1403)

Dill and Thill (2007) used two coders and a careful training procedure for their analysis of the magazine images about video games:

> One male and one female rater, both undergraduate psychology majors, practiced on images from magazines similar to those used in the current investigation. Raters discussed these practice ratings with each other and with the first author until they showed evidence of properly applying the coding scheme for all variables. Progress was also checked part way through the coding process, as suggested by Cowan (2002). Specifically, the coding scheme was re-taught by the first author, and the two raters privately discussed discrepancies and then independently assessed their judgments about their ratings of the discrepant items. They did not resolve discrepancies, but simply reconsidered their own ratings in light of the coding scheme refresher session. Cowan (2002) reports that this practice of reevaluating ratings criteria is of particular value when coding large amounts of violent and sexual material because, as with viewers, coders suffer from desensitization effects. (p. 856)

Ambwani and Strauss (2007) also designed a careful training process to achieve acceptable levels of reliability before their raters coded the written answers to their open-ended survey questions:

> We developed a coding scheme derived from common themes in the qualitative responses; descriptions of these themes appear in the Appendix. We independently developed lists of possible coding themes by reviewing the responses and then came to a consensus regarding the most frequently emerging themes. Next, a team of four independent raters was trained in the coding scheme through oral, written, and group instruction by the first author. The raters first coded a set of hypothetical responses; once the raters reached an acceptable level of agreement on the coding for the hypothetical responses, they began to code the actual data. These strategies, recommended by Orwin (1994), were employed to reduce error in the data coding. All four independent coders were blind to the hypotheses. Each qualitative response could be coded for multiple themes. Thus, for example, a participant who described the influence of body image in mate selection and quality of sexual relations received positive codes for both *choice* and *quality of sex.* (p. 16)

Gerth and Siegert (2012) do not provide details about their coding procedures.

Develop Appropriate Statistical Analyses

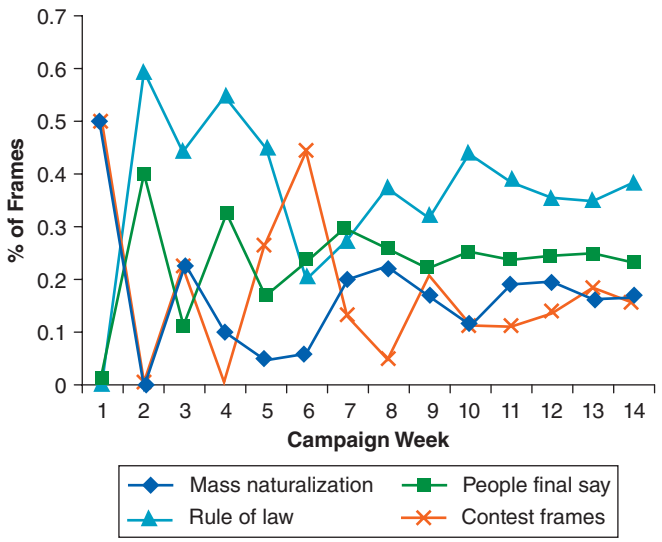
The content analyst creates variables for analysis by counting occurrences of particular words, themes, or phrases and then tests relations between the resulting variables. These analyses could use some of the statistics introduced in Chapter 9, including frequency distributions, measures of central tendency and variation, cross-tabulations, and correlation analysis (Weber 1990:58–63). Computer-aided qualitative analysis programs, like those you learned about in Chapter 11 and like the one I selected for the preceding newspaper article analysis, can help, in many cases, to develop coding procedures and then carry out the content coding.

The simple chart that Neuendorf (2002:3) used to analyze the frequency of medical programming appears in Figure 15.12. My content analysis with Duckworth and others (2003) was simply a comparison of percentages showing that 28% of the articles mentioning schizophrenia used it as a metaphor, compared with only 1% of the articles mentioning cancer. We also presented examples of the text that had been coded into different categories. For example, *the nation's schizophrenic perspective on drugs* was the type of phrase coded as a metaphorical use of the term *schizophrenia* (p. 1403). Dill and Thill (2007:858) presented percentages and other statistics that showed that, among other differences, female characters were much more likely to be portrayed in sexualized ways in video game images than were male characters. Ambwani and Strauss (2007:16) used other statistics that showed that body esteem and romantic love experiences are related, particularly for women. They also examined the original written comments and found further evidence for this relationship. For example, one woman wrote, "[My current boyfriend] taught me to love my body. Now I see myself through his eyes, and I feel beautiful" (p. 17).

Gerth and Siegert (2012:288–295) use both charts and percentage distributions to test the hypotheses they posed about media attention to different political perspectives during political campaigns. Figure 15.12 shows how the use of different perspectives (or "frames") varied over the weeks of the campaign, with an emphasis on the "rule of law" being the most common way of framing the issue of naturalization.

The criteria for judging quantitative content analyses of text are the same standards of validity applied to data collected with other quantitative methods. We must review the sampling approach, the reliability and validity of the measures, and the controls used to strengthen any causal conclusions.

FIGURE 15.12 ■ Frames per Campaign Week



Source: Gerth, Matthias A., and Gabriele Siegert. 2012. "Patterns of Consistence and Constriction: How News Media Frame the Coverage of Direct Democratic Campaigns." *American Behavioral Scientist* 56:279–299.

ETHICAL ISSUES IN HISTORICAL AND COMPARATIVE RESEARCH AND CONTENT ANALYSIS

Analysis of historical documents, documents from other countries, or content in media does not create the potential for harm to human subjects that can be a concern when collecting primary data. It is still important to be honest and responsible in working out arrangements for data access when data must be obtained from designated officials or data archivists, but many data are available easily in libraries or on the web. Researchers in the United States who conclude that they are being denied access to public records of the federal government may be able to obtain the data by filing a Freedom of Information Act (FOIA) request. The FOIA stipulates that all persons have a right to access all federal agency records unless the records are specifically exempted (Riedel 2000:130–131). Researchers who review historical or government documents must also try to avoid embarrassing or otherwise harming named individuals or their descendants by disclosing sensitive information.

Ethical concerns are multiplied when surveys are conducted or data are collected in other countries. If the outside researcher lacks much knowledge of local norms, values, and routine activities, the potential for inadvertently harming subjects is substantial. For this reason, cross-cultural researchers should spend time learning about each of the countries in which they plan to collect primary data and strike up collaborations with researchers in those countries (Hantrais and Mangen 1996). Local advisory groups may also be formed in each country so that a broader range of opinion is solicited when key decisions must be made. Such collaboration can also be invaluable when designing instruments, collecting data, and interpreting results.

Cross-cultural researchers who use data from other societies have a particular obligation to try to understand the culture and norms of those societies before they begin secondary data analyses. It is a mistake to assume that questions asked in other languages or cultural contexts will have the same meaning as when asked in the researcher's own language and culture, so a careful, culturally sensitive process of review by knowledgeable experts must precede measurement decisions in these projects. Ethical standards themselves may vary between nations and cultures, so cross-cultural researchers should consider collaborating with others in the places to be compared and take the time to learn about cultural practices, gender norms, and ethical standards (Ayhan 2001; Stake and Rizvi 2009:527).

CONCLUSIONS

Historical and comparative social science investigations use a variety of techniques that range from narrative histories having much in common with qualitative methods to analyses of secondary data that are in many respects like other quantitative analyses of survey data. Content analysis may also be used in historical and comparative research. Demographic analysis can enrich other forms of historical and comparative research by describing differences in population structure and dynamics between countries and over time. Each of these techniques can help the researchers gain new insights into processes such as democratization. They encourage intimate familiarity with the cause of development of the nations studied and thereby stimulate inductive reasoning about the interrelations between different historical events. Systematic historical and comparative techniques can be used to test deductive hypotheses concerning international differences as well as historical events.

Most historical and comparative methods encourage causal reasoning. They require the researcher to consider systematically the causal mechanism, or historical sequences of events, by

which earlier events influence later outcomes. They also encourage attention to causal context, with a particular focus on the ways in which different cultures and social structures may result in different effects of other variables. Content analysis methods focus attention on key dimensions of variation in news media and other communicative texts, while demographic analysis adds another important dimension. There is much to be gained by learning and continuing to use and develop these methods.

KEY TERMS

Case-oriented research
Comparative historical research (p. 524)
Conjunctural research
Content analysis (p. 545)
Cross-sectional comparative research (p. 524)
Demography (p. 543)
Event-structure analysis (p. 526)
Historical events research (p. 524)
Historical process research (p. 524)
Holistic research
Method of agreement (p. 540)
Method of difference (p. 541)
Narrative explanation
Oral history
Temporal research
Variable-oriented research (p. 531)

HIGHLIGHTS

LO 15.1 Appraise the value of historical and comparative methods with one historical or contemporary example.

- The central insight behind historical and comparative methods is that we can improve our understanding of social processes when we make comparisons with other times and places.
- Max Weber and Emile Durkheim both used historical and comparative methods in laying the foundation for the discipline of sociology.
- Casanova's research on the impact of adjustment programs on political culture in Europe is a contemporary example of the value of these methods.
- Methodological challenges for comparative and historical research include missing data, variation in the meaning of words and phrases and in the boundaries of geographic units across historical periods and between cultures, bias or inaccuracy of historical documents, lack of measurement equivalence, the need for detailed knowledge of the cases chosen, a limited number of cases, case selection based on availability, reliance on dichotomous categorization of cases, and interdependence of cases selected.

LO 15.2 Contrast the two types of historical social science methods.

- Historical events research and historical process research are both often qualitative.
- Qualitative historical research is case oriented, holistic, conjunctural, temporal, historically specific, narrative (idiographic), and inductive.
- Event-structure analysis is a systematic qualitative approach to developing an idiographic causal explanation for a key event.
- Oral histories use data collected through intensive interviews with participants to describe past events.

- Historical events research focuses attention on a significant social event at one time in the past, while historical process research studies processes over a long period of time in the past.
- Qualitative historical process research uses a narrative approach to causal explanation, in which historical events are treated as part of a developing story. Narrative explanations are temporal, holistic, and conjunctural.

LO 15.3 Contrast the three types of comparative social science methods.

- Comparative methods may be cross sectional, such as when variation between country characteristics is compared, or historical, in which developmental patterns are compared between countries.
- Cross-sectional comparative research and comparative historical research are both often quantitative.
- Cross-sectional comparative research compares data from one period between two or more nations or other areas.
- Comparative historical research compares data from more than one period in more than one nation or area.
- Comparative case study designs use the method of agreement and/or the method of difference to identify variables that make a difference for an outcome.

LO 15.4 Appraise the value of adding demographic analysis to historical and comparative investigations.

- The social processes that are the focus of historical and comparative research are often reflected in and influenced by changes in the makeup of the population being studied.
- Demography is the statistical and mathematical study of the size, composition, and spatial distribution of human populations and how these features change over time.
- Central concepts for demographic research are population change, standardization of population numbers, the demographic bookkeeping equation, and population composition.
- Demographers explain population change through five processes: (1) fertility, (2) mortality, (3) marriage, (4) migration, and (5) social mobility.

LO 15.5 Critique content analysis as a way of identifying social patterns.

- Content analysis is a tool for systematic quantitative analysis of documents and other textual data.
- Content analysis requires careful testing and control of coding procedures to achieve reliable measures.
- The quantitative counts in content analysis can overlook the meaning of arguments in passages of text.

LO 15.6 Formulate an argument for exempting historical and comparative research from IRB review.

- Analysis of historical documents or secondary data does not create new potential harm for human subjects.
- The Freedom of Information Act (FOIA) stipulates that all persons have a right to access all federal agency records unless records are specifically exempted.
- Ethical historical and comparative research procedures avoid embarrassing or otherwise harming named individuals or their descendants.
- Ethical cross-cultural research respects the norms of other cultures.

DISCUSSION QUESTIONS

1. Review the differences between case-oriented, historically specific, inductive explanations and those that are more variable-oriented, theoretically general, and deductive. List several arguments for and against each approach. Which is more appealing to you and why?
2. Consider designing a historical comparative study of influences on the growth of support for authoritarianism in the political cultures of the United States and Europe. Would you expect to find similar processes in the United States to those identified by Casanova in Europe? What other possible influences do you think should be compared between the two?
3. How has the coronavirus pandemic influenced social patterns in the nation? What evidence would you suggest collecting for a social science investigation of the pandemic as a historical event?
4. Consider the comparative historical research by Rueschemeyer et al. (1992) on democratic politics in Latin America. What does comparison between nations add to the researcher's ability to develop causal explanations?
5. Kathleen Fallon, Liam Swiss, and Jocelyn Vitema (2012) developed a *nomothetic* causal explanation of variation in the representation of women in national legislatures over time, whereas Griffin's (1993) explanation of a lynching can be termed *idiographic*. Discuss the similarities and differences between these types of causal explanations. Use these two studies to illustrate the strengths and weaknesses of each.
6. Select a major historical event or process, such as the civil rights movement, the Vietnam War, Brexit, or the invasion of Ukraine. Why do you think this event happened? Now, select one of the four major types of historical and comparative methods that you think could be used to test your explanation. Why did you choose this method? What type of evidence would support your proposed explanation? What problems might you face in using this method to test your explanation?
7. Consider the media that you pay attention to in your social world. How could you design a content analysis of the messages conveyed by these media? What research questions could you help to answer by adding a comparison to another region or country as part of this content analysis?

SPSS EXERCISES

1. In this exercise, you will conduct a cross-sectional comparative analysis of regional differences in political culture. For this cross-sectional comparative analysis, you will compare several attitudes measured in the GSS across the four broad regions of the United States.

a. Create crosstabs that display the association between religious, political, and social attitudes and region of residence. After opening the GSS2020 dataset, from the main SPSS menu choose Analyze/Descriptive Statistics/Crosstabs. Choose some or all of the following variables as the row variables: trustsci_1b, affrmact_1b, wrkwayup_1b marhomo1_1b fefam_1b fejobaff_1b, fund2020, trump2020.

b. Choose region4 as the column variable and request Cells=percentages=column.

c. Run the analysis and inspect the output.

d. How do attitudes of Americans differ between the four regions? Describe the similarities and differences you have found.

©iStockphoto.com/ridvan_celik

16 SUMMARIZING AND REPORTING RESEARCH

LEARNING OBJECTIVES

Upon completion of this chapter, the reader should be able to

16.1 Critique your own approach to writing papers and reports in relation to the principles formulated by experienced writers

16.2 Illustrate several types of charts used to display research

16.3 Argue for using meta-analysis as a tool for advancing science

16.4 Compare the advantages of reporting research through journal articles, applied research reports, research videos, and posters

16.5 Discuss examples of following ethical guidelines in an academic journal article and of failing to follow ethical guidelines

16.6 Appraise the motivations for plagiarism and the ways of minimizing them

You learned in Chapter 2 that research is a circular process, so it is appropriate that we end this book where we began. The stage of reporting research results is also the point at which the need for new research is identified. It is the time when, so to speak, "the rubber hits the road"—when we must make our research make sense to others. To whom will our research be addressed? How should we present our results to them? Should we seek to influence how our research report is used? These are just a few of the questions to consider at this critical point.

The primary goals of this chapter are to guide you in writing worthwhile reports of your own, displaying findings, and communicating with the public about research. This chapter gives particular attention to the writing process itself and points out how that process can differ when writing qualitative versus quantitative research reports. I will conclude by considering some of the ethical issues unique to the reporting process, with special attention to the problem of plagiarism.

WRITING RESEARCH

The goal of research is not just to discover something but also to communicate that discovery to a larger audience: other social scientists, government officials, your teachers, the general public—perhaps several of these audiences. Whatever the study's particular outcome, if the intended audience for the research comprehends the results and learns from them, the research can be judged a success. If the intended audience does not learn about the study's results, the research should be judged a failure—no matter how expensive the research, how sophisticated its design, or how much you (or others) invested in it.

Successful research reporting requires both good writing and a proper publication outlet. We will first review guidelines for successful writing before we look at particular types of research publications.

Consider the following principles formulated by experienced writers (Booth, Colomb, and Williams 1995:150–151):

- Respect the complexity of the task and don't expect to write a polished draft in a linear fashion. Your thinking will develop as you write, causing you to reorganize and rewrite.

- Leave enough time for dead ends, restarts, revisions, and so on, and accept the fact that you will discard much of what you write.
- Write as fast as you comfortably can. Don't worry about spelling, grammar, and so on until you are polishing things up.
- Ask anyone whom you trust for reactions to what you have written.
- Write as you go along, so that you have notes and report segments drafted even before you focus on writing the report.

It is important to outline a report before writing it, but neither the report's organization nor the first written draft should be considered fixed. As you write, you will get new ideas about how to organize the report. Try them out. As you review the first draft, you will see many ways to improve your writing. Focus particularly on how to shorten and clarify your statements. Make sure that each paragraph concerns only one topic. Remember the golden rule of good writing: Writing is revising!

You can ease the burden of report writing in several ways:

- Draw on the research proposal and on project notes.
- Use a word processing program on a computer to facilitate reorganizing and editing.
- Seek criticism from friends, teachers, and other research consumers before turning in the final product.

I find it helpful at times to use what I call **reverse outlining**: After you have written a first complete draft, outline it on a paragraph-by-paragraph basis, ignoring the section headings you used. See if the paper you wrote fits the outline you planned. Consider how the organization could be improved.

Most important, leave yourself enough time so that you can revise, several times if possible, before turning in the final draft. Here are one student's reflections on writing and revising:

> I found the process of writing and revising my paper longer than I expected. I think it was something I was doing for the first time—working within a committee—that made the process not easy. The overall experience was very good, since I found that I have learned so much. My personal computer also did help greatly.
>
> Revision is essential until complete clarity is achieved. This took most of my time. Because I was so close to the subject matter, it became essential for me to ask for assistance in achieving clarity. My committee members, English editor, and fellow students were extremely helpful. Putting it on disk was also, without question, a timesaver. Time was the major problem.
>
> The process was long, hard, and time-consuming, but it was a great learning experience. I work full time so I learned how to budget my time. I still use my time productively and am very careful of not wasting it. (Graduate Program in Applied Sociology 1990:13)

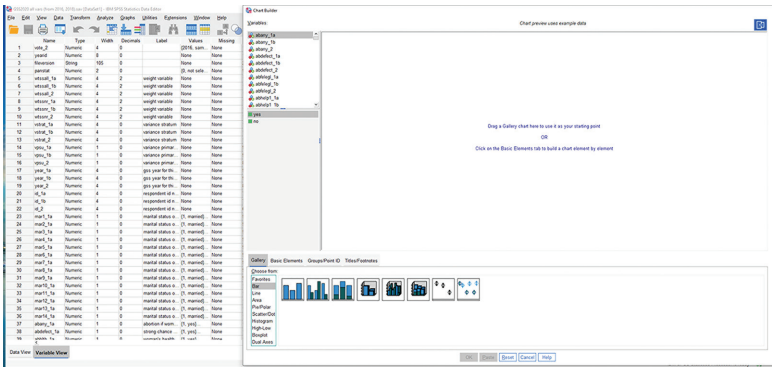
For more suggestions about writing, see Howard Becker (2007), Booth, Colomb, and Williams (2008), Lee Cuba (2002), William Strunk Jr. and E. B. White (2000), and Kate Turabian (2007).

DISPLAYING RESEARCH

Graphic displays will help draw attention to important aspects of your findings and simplify interpretation of basic patterns. You should use graphic displays judiciously in a report, but use them you must if you want your findings to be accessible to the widest possible audience. Chapter 9 introduced some conventions for reporting quantitative research results, including tabular displays and graphs. You can improve your displays of quantitative data in research reports by using the *Chart Builder* feature of SPSS (Aldrich and Rodriguez 2013).

Open the Chart Builder by clicking on "Graphs" and then "Chart Builder" in the SPSS Data Editor screen, with the GSS2020 data set. Figure 16.1 shows the SPSS Data Editor window as it appears at that point.

FIGURE 16.1 ■ Building a Chart With Chart Builder



To build a simple bar chart, click the "Gallery" option and then "Bar" in the "Choose from" list and drag the picture of the bar chart to the "Chart Preview" window. Now drag and drop the categorical variable whose distribution you would like to display (try "partyid_2") into the "X-Axis?" label box. If you would like to alter the graph or additional elements, click on the "Element Properties" tab and make desired changes. A good idea is to change the frequencies ("count") to percentages: Change the Bar 1 label to Percent and click "Apply." Click "OK" to generate the chart. See Figure 16.2.

FIGURE 16.2 ■ Adding Chart Features With Properties

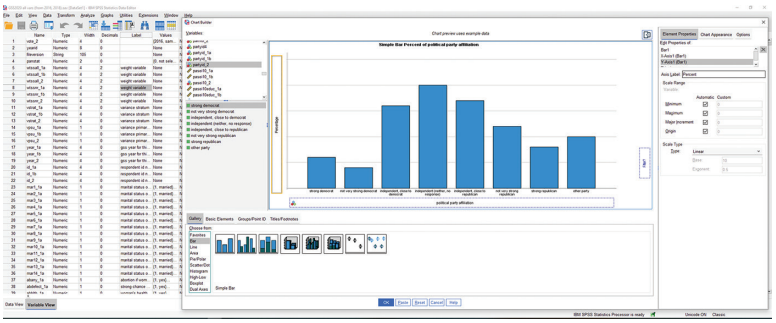
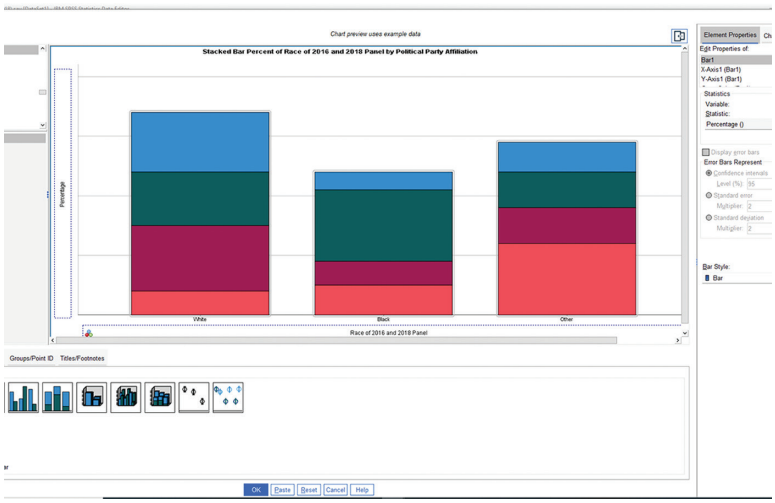


Chart Builder can also be used to create figures that display the association between variables. Figure 16.3 shows how to use Chart Builder to create a stacked bar chart to display the association between party identification recoded to four categories (partyid4) and race dichotomized (raced) with General Social Survey (GSS) 2020 data. To create this chart, click "Reset" at the

bottom of the Chart Builder window in order to begin a new chart. Then click the "Stacked Bar" icon from the Gallery and drag it into the chart builder window. Under Variables, select "raced" and drag it onto the X-Axis? heading for the bars. Now drag "partyid4" onto the "Stack: set color" heading in the upper-right corner. On the right side, Edit Properties of: "Bar1" and, under Statistics, change "Count" to "Percentages (?)." Then, also on the right, choose "Set Parameters" and select as the "Denominator for Computing Percentages": "Total for Each X-Axis Category." Now your screen should look like the one in Figure 16.3

FIGURE 16.3 ■ Creating a Stacked Bar Display of a Bivariate Association With Chart Builder



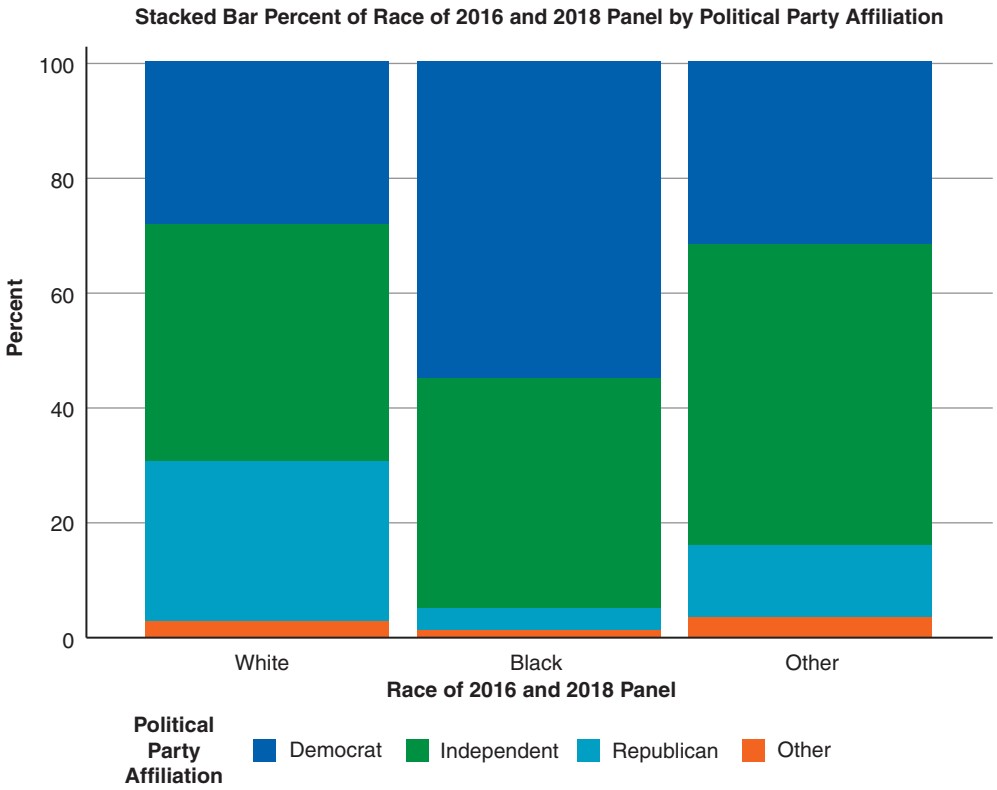
Click "OK." Figure 16.4 shows the resulting stacked bar chart displaying the association between party identification recoded to four categories (partyid4) and the three categories of race (race2020) recorded in the General Social Survey (GSS) 2020 data. You can see at a glance that white respondents were much more likely to identify as Republicans (almost 30%) than Black respondents (about 5%) or those identifying with another race (about 10%). By contrast, Black respondents (more than half) were much more likely to identify themselves as Democrats than either white (just over 25%) or "Other" respondents (about 30%).

Graphic displays will help draw attention to important aspects of your findings and simplify the description of basic patterns. You should use graphic displays judiciously in a report, but use them you must if you want your findings to be accessible to the widest possible audience.

Data visualization is a rapidly developing field that focuses on how to represent and present data to facilitate understanding (Kirk 2016:19). It goes a step further than just picking a chart type from a PowerPoint menu or statistical package like SPSS. By considering carefully how each element of a data presentation can be used to maximize understanding of what the presentation contains, data visualization can convey more information that can be easier to comprehend. Understanding is most likely to be achieved in a data display that is

- *Trustworthy.* The basic pattern is not distorted to change the viewer's impression. For example, adding a three-dimensional appearance to figures in a chart can make them appear larger than they actually are.
- *Accessible.* It should be easy for the viewer to see the key elements in the data and the relationships between them. Decisions about how simple a display needs to be have to

FIGURE 16.4 ■ Creating a Stacked Bar Display of a Bivariate Association With Chart Builder

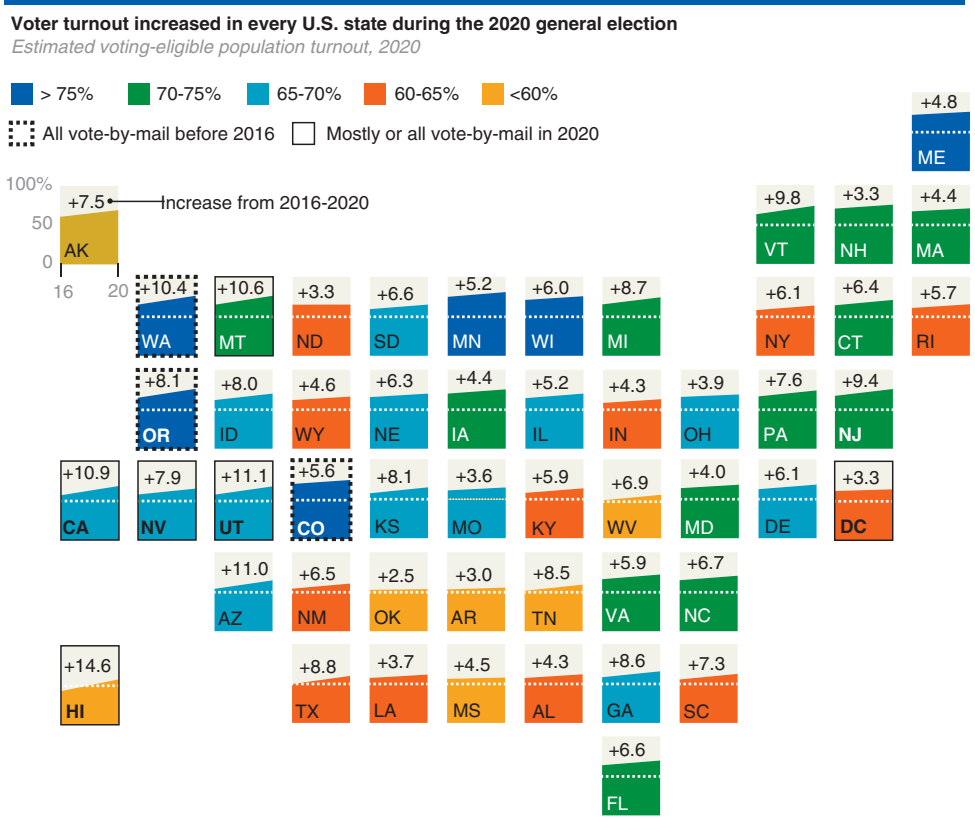


take into account the expertise of the intended audience and the time they will have to view the display.

- *Elegant.* The design should attract and sustain attention. It's easy to overdo it, but try to make the design engaging as well as informative.

Figure 16.5 shows how attention to visualization in a chart can help to convey information. The squares make it easy to compare variation in voter turnout between states, while displaying the states with a map of the United States makes it easy for those familiar with the nation to see which state is which and to readily see how regions compare—and perhaps how their own state compares to others. The trend lines in the boxes show how turnout changed between the presidential elections in 2016 and 2020, while the numbers above the trend lines identify the specific amount of change. In addition, the color scheme (with labels in the legend above the figure) provides a quick impression of the total state voter turnout in 2020—reinforcing the information also conveyed by the proportion of each box that is filled in with the corresponding color. Finally, the use of white and black letters for the state abbreviations and of solid or broken lines, or no lines to frame the boxes distinguishes states using vote-by-mail exclusively before 2016, those using it mostly in 2020, and those not using vote-by-mail. That's a lot of information—much more than a simple bar chart can convey (see Chapter 9), but the careful visualization makes it easier to answer questions about variation in voting between states and how it changed from 2016 to 2020, and even why it might have changed (allowing vote-by-mail). See Andy Kirk's (2016) *Data Visualisation* book from SAGE for many other examples and check out the book's website, http://book.visualisingdata.com/home.

FIGURE 16.5 ■ Increase in Voter Turnout in 2020 Presidential Election, by Age



Source: DeSilver, Drew. 2021. *Turnout Soared in 2020 as Nearly Two-Thirds of Eligible U.S. Voters Cast Ballots for President*. Washington, DC: Pew Research Center, January 28.

PERFORMING META-ANALYSES

Meta-analysis is a quantitative method for identifying patterns in findings across multiple studies of the same research question (Cooper and Hedges 1994). Unlike a traditional literature review, which describes previous research studies verbally, meta-analyses treat previous studies as cases, whose features are measured as variables and then analyzed statistically. It is kind of like conducting a survey in which the "respondents" are previous studies. Meta-analysis shows how evidence about social processes varies across research studies. If the methods used in these studies varied, then meta-analysis can describe how this variation affected the study findings. If social contexts varied across the studies, then meta-analysis will indicate how social context affected the study findings. You have already learned in this chapter about most of the statistics that are used in meta-analyses.

Meta-analysis can be used when a number of studies have attempted to answer the same research question with similar quantitative methods, most often experiments. Meta-analysis is not appropriate for evaluating results from qualitative studies or from multiple studies that used different methods or measured different dependent variables. It is also not very sensible to use meta-analysis to combine study results when the original case data from these studies are available and can be combined and analyzed together (Lipsey and Wilson 2001). Meta-analysis

is a technique for combination and statistical analysis of published research reports, although meta-analysts may contact the original researchers to add information about studies that was not included in the published reports..

After a research problem is formulated based on the findings of prior research, the literature must be searched systematically to identify the entire population of relevant studies (see Chapter 2). Typically, multiple bibliographic databases are used; some researchers also search for relevant dissertations and conference papers. Once the studies are identified, their findings, methods, and other features are coded (e.g., sample size, location of sample, and strength of the association between the independent and dependent variables). Eligibility criteria must be specified carefully to determine which studies to include and which to omit as too different. (Mark Lipsey and David Wilson 2001:16–21) suggested that eligibility criteria include the following:

- *Distinguishing features.* This includes the specific intervention tested and perhaps the groups compared.
- *Research respondents.* The pertinent characteristics of the research respondents (subject sample) who provided study data must be similar to those of the population about which generalization is sought.
- *Key variables.* These must be sufficient to allow tests of the hypotheses of concern and controls for likely additional influences.
- *Research methods.* Apples and oranges cannot be directly compared, but some trade-off must be made between including the range of studies about a research question and excluding those that are so different in their methods as not to yield comparable data.
- *Cultural and linguistic range.* If the study population is going to be limited to English-language publications or limited in some other way, this must be acknowledged, and the size of the population of relevant studies in other languages should be estimated.
- *Time frame.* Social processes relevant to the research question may have changed for reasons such as historical events or the advent of new technologies, so temporal boundaries around the study population must be considered.
- *Publication type.* It must be determined whether the analysis will focus only on published reports in professional journals or include dissertations and unpublished reports.

Statistics are then calculated to identify the average effect of the independent variable on the dependent variable, as well as the effect of methodological and other features of the studies (Cooper and Hedges 1994). The **effect size** statistic is the key to capturing the association between the independent and dependent variables in a way that is comparable across multiple studies. The effect size statistic is a standardized measure of association—often the difference between the mean of the experimental group and the mean of the control group on the dependent variable, adjusted for the average variability in the two groups (Lipsey and Wilson 2001).

The meta-analytic approach to synthesizing research findings can result in much more generalizable findings than those obtained with just one study. Methodological weaknesses in the studies included in the meta-analysis are still a problem, however; only when other studies without major methodological weaknesses are included can we estimate effects with some confidence. In addition, before we can place any confidence in the results of a meta-analysis, we must

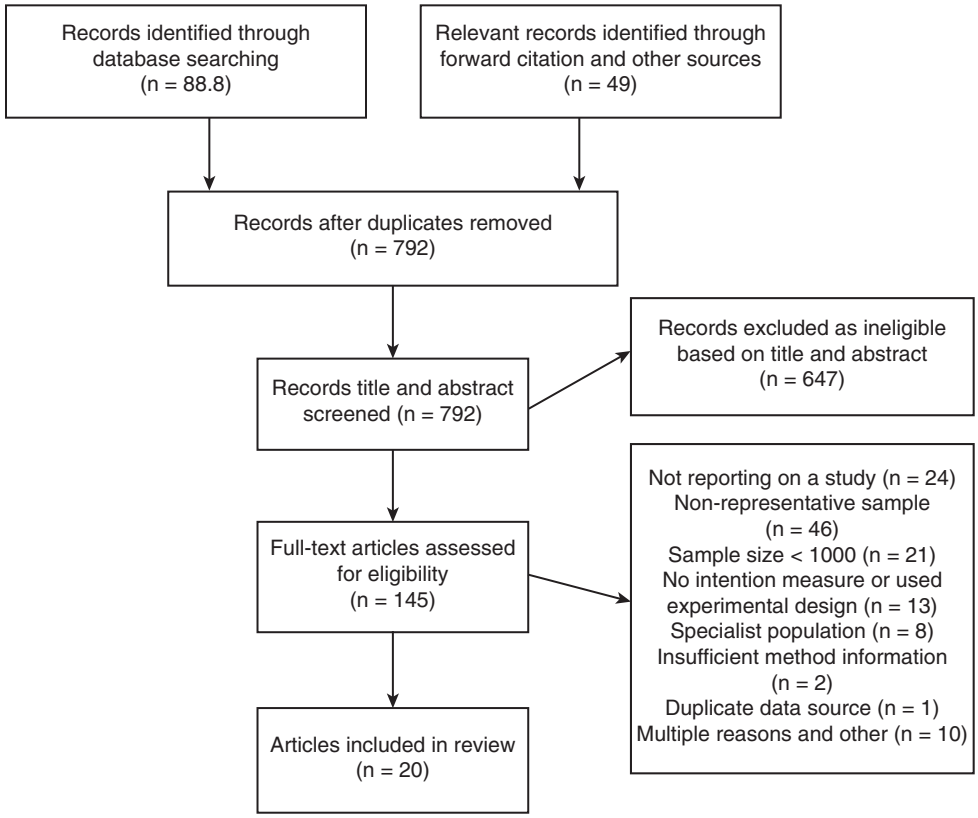
be confident that all (or almost all) relevant studies were included and that the information we need to analyze was included in all (or most) of the studies, or can be obtained directly from the original researchers (Matt and Cook 1994).

Case Study: Vaccine Intentions Across 28 Large National Samples

Did COVID-19 vaccination intentions vary across countries? Did they change over time in the pandemic's first year? Eric Robinson, Andrew Jones, and India Lesser at the University of Liverpool in England and Michael Daly at Maynooth University in County Kildare, Ireland sought to answer these questions with a meta-analysis of published research. Their research report illustrates the key steps in a meta-analysis (Robinson et al. 2021).

Robinson et al. (2021) began their analysis with a comprehensive review of studies that measured intentions or willingness to be vaccinated against the COVID-19 virus in nationally representative samples that were either published or awaiting publication after acceptance in a peer-reviewed journal. Only studies with a sample size of 1,000 or more members of the general population were eligible, and they had to be included in at least one of three bibliographic databases: PUBMED, Scopus, or three preprint (before publication) archives. This search identified 145 full-text articles, but only 20 met all study criteria and represented general population surveys of national populations (see Figure 16.6). The 20 articles represented 28 samples of a total of 58,656 participants in a total of 13 countries.

FIGURE 16.6 ■ Study Selection Flowchart


Records identified through database searching (n = 88.8)
Relevant records identified through forward citation and other sources (n = 49)
Records after duplicates removed (n = 792)
Records title and abstract screened (n = 792)
Records excluded as ineligible based on title and abstract (n = 647)
Full-text articles assessed for eligibility (n = 145)
Not reporting on a study (n = 24)
Non-representative sample (n = 46)
Sample size < 1000 (n = 21)
No intention measure or used experimental design (n = 13)
Specialist population (n = 8)
Insufficient method information (n = 2)
Duplicate data source (n = 1)
Multiple reasons and other (n = 10)
Articles included in review (n = 20)

Source: Robinson et al. (2021:2027). Vaccine.

Robinson and his coauthors then summarized the characteristics of each study (see Figure 16.7) and, in a separate step, each of the study's major findings.

FIGURE 16.7 ■ Summary of Studies in Meta-Analysis*

Author	Country	Journal article or pre-print	Probability vs. quota sampling[1]	N	Month	Measure[2]	Intention results[3]
Loomba	UK	Pre-print	Quota	4001	September	If a new coronavirus (COVID-19) vaccine became available, would you accept the vaccine for yourself? *Yes [Y], definitely [Y], unsure but leaning towards yes [U], unsure but leaning towards no [U]. no definitely not [N]*	Yes: 54% Unsure: 40% No: 6%
Romer	US	Journal article	Probability	1050	March	If there were a vaccine that protected you from getting the coronavirus, how likely, if at all. would you be to decide to be vaccinated? *Not at all likely [N], not too likely [N], somewhat likely [Y], very likely [Y]*	Yes: 86% No: 15%
Taylor	US	Journal article	Quota	1772	May	If a vaccine for COVID-19 was available, would you get vaccinated? *Yes [Y], no [N]*	Yes: 75% No: 25%
Carpiano	US	Pre-print	Probability	1000	May	If a vaccine against the coronavirus becomes available, do you plan to get vaccinated, or not? *Yes [Y], no [N], not sure [U]*	Yes: 50% Unsure: 30% No: 20%
Callaghan	US	Pre-print	Quota	5009	May-June	Scientists around the world are working on developing a vaccine to protect individuals against the coronavirus. If a Vaccine is developed, would you pursue getting vaccinated for the coronavirus? *Yes [Y], no [N]*	Yes: 69% No: 31%
McAndrew	US	Pre-print	Quota	1198	June	When a Coronavirus (COVID-19) vaccine becomes available,do you think you will or will not get vaccinated? *Definitely will get vaccinated [Y], probably will get vaccinated [Y], probably will not get vaccinated [N], definitely will not get vaccinated [N], unsure [U]*	Yes: 59% Unsure: 15% No: 25%
Loomba	US	Pre-print	Quota	4000	September	If a new coronavirus (COVID-19) vaccine became available, would you accept the vaccine for yourself? *Yes definitely [Y], unsure but leaning towards yes [U], unsure but leaning towards no [U], no definitely not [N]*	Yes: 44% Unsure: 41% No: 15%
Daly	US	Pre-print	Probability	7547	October	How likely are you to get vaccinated for coronavirus when a vaccine becomes available *Unsure [U], somewhat unlikely [N], very unlikely [N], somewhat likely [Y], very likely [Y]*	Yes: 54% Unsure: 14% No: 32%

[1] If sampling method was unclear from the description in the study method, we confirmed use of quota vs. probability by searching for the data source online (e.g. panel provider website).

[2] [Y], [N] and [U] indicate extracted response options representing yes, no and unsure in the present meta-analysis.

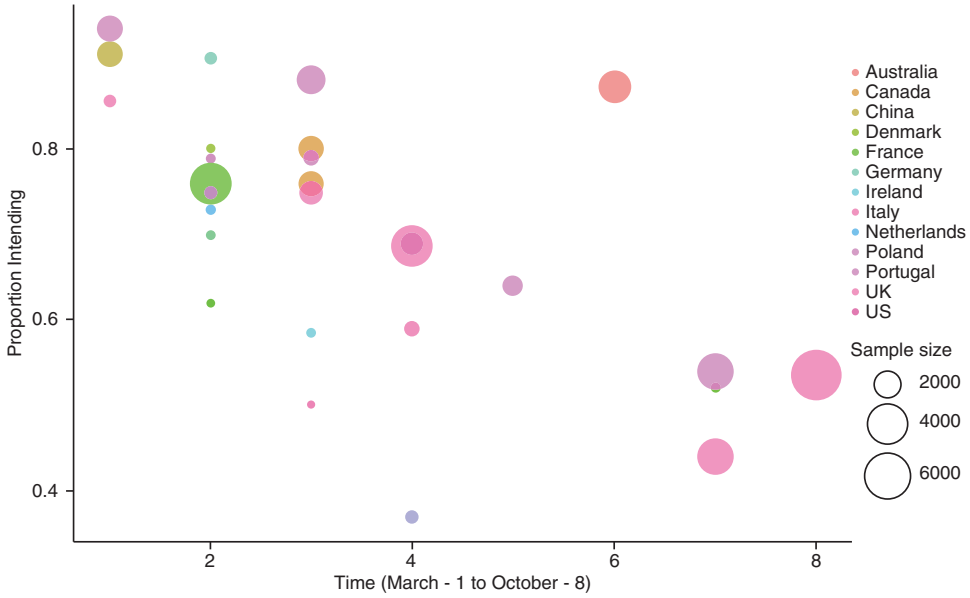
[3] Values may not equal 100 due to rounding of reported values in study manuscripts.

*Partial table

Source: Robinson et al. (2021:2029). Vaccine.

Finally, each study was classified according to the time of the study (from March to October 2020) and the proportion found to be intending to be vaccinated. Robinson and his coauthors concluded that on average, about 73% intended to be vaccinated but that this intention had declined over time (see Figure 16.8). (Bear in mind that this decline occurred before COVID-19 vaccinations first became available in December 2020). They also found that likelihood of intending to be vaccinated was greater among men, respondents who were older, those who were white, and among those with higher income or education.

FIGURE 16.8 ■ Summary of Outcomes in Meta-Analysis



Source: Robinson et al. (2021:2031). Vaccine.

Can you see why the systematic literature reviews you learned about in Chapter 2 often include a statistical meta-analysis of the findings from the studies reviewed? It is an excellent way to summarize the central findings in prior research and to highlight the extent to which these findings have been consistent. A meta-analysis also reminds us that the results of a single study must be interpreted in relation to the larger body of research findings.

REPORTING RESEARCH

You begin writing your research report when you are working on the research proposal and writing your literature review (see Chapter 2). You will find the final report is much easier to write, and more adequate, if you write more material for it as you work out issues during the project. It is very disappointing to discover that something important was left out when it is too late to do anything about it. Students and professional researchers alike frequently leave final papers and reports until the last possible minute, often for understandable reasons, including other coursework and job or family responsibilities. But be forewarned: *Waiting until the last minute to write the report will not result in the best possible final product.*

The organization of your research report will depend to some extent on the audience for which you are writing and the type of research you have conducted. Articles that will be submitted to an academic journal will differ from research reports written for a funding agency or for the general public.

Research reports based on qualitative research will differ in some ways from those based on quantitative research. Students writing course papers are often required to structure their research reports using the journal article format, and they may be asked to present their results differently if they have used qualitative (or mixed) methods. The following sections outline the major issues to consider.

Journal Articles

Writing for academic journals is perhaps the toughest form of writing because articles are submitted to several experts in your field for careful review—anonymously, with most journals—before acceptance for publication. Perhaps it wouldn't be such an arduous process if so many academic journals did not have rejection rates in excess of 90% and turnaround times for reviews that are usually several months. Even the best articles, in the judgment of the reviewers, are most often given a "revise and resubmit" after the first review and then are evaluated all over again after the revisions are concluded.

But journal article review procedures have some important benefits. Most important is the identification of areas in need of improvement, as the eyes of the author(s) are replaced by those of previously uninvolved subject-matter experts and methodologists. A good journal editor makes sure they have a list of many different types of experts available for reviewing whatever types of articles the journal is likely to receive. There is a parallel benefit for the author(s): It is always beneficial to review criticisms of your own work by people who know the field well. It can be a painful and time-consuming process, but the entire field moves forward as researchers continually critique and suggest improvements in each other's research reports.

Are you familiar with the expression "no pain, no gain"? It applies here, too. Hundreds of online journals masquerade as scholarly journals but are simply money-making ventures for their "publishers" (Anderson 2017). These journals solicit articles from people eager to publish their papers, who then may pay thousands of dollars to have their work included in an upcoming issue of the "open access" online journal. There is no peer-review process and often no real copyediting, so articles can be published with bizarre claims and ungrammatical prose. Studies of journal practices involving submitting phony (and absurd) articles have found that such "predatory journals" will accept anything—for a fee (Bohannon 2013). They also recruit widely for editors, and research shows they will appoint anyone to their editorial boards without regard to qualifications (Sorowski et al. 2017).

The problem is not with open access journals *per se*, for there are now many highly respected open access journals, and most print journals now also offer an on-line version of their journal (as you should know from your literature searches) and allow authors who pay a hefty fee to have the full text of their articles available to all users (who are not so fortunate to have access through their university library). The problem is with open access journals that do not use a professional peer-review process and are just seeking revenue.

Case Study: Does Protest Against Police Violence Matter?

Figure 16.9 presents an outline of the sections in an academic journal article, with some illustrative quoted text. In the article, Susan Olzak tests hypotheses about the effect of protests against police violence. She begins with a clear abstract of the article, which summarizes in one paragraph the research question, prior research, study methods and major findings, and key conclusions. Many others who search the literature about the topic will never read an entire article unless they are persuaded by the abstract that the article provides worthwhile and relevant information, so the abstract should be written and edited carefully. The article's introduction highlights the problem selected—in the first figure in the article, that's the relationship between protests against police violence and change in police policies and use of lethal force—and clearly identifies its importance: the need to understand the role of social movements in influencing change in a democratic system. Literature

reviews in journal articles should be integrated reviews that highlight major relevant findings and identify key methodological lessons from the prior research as a whole rather than presenting separate summaries of prior research studies (see Chapter 2). Olzak does this by reviewing the literature in three sections—Does Protest Matter? Changes in Policing and Police Violence, Community Response: Demanding Civilian Oversight—and then outlining her core argument in a Theoretical Argument section that begins with the first figure in her article and ends with three hypotheses.

The article's Research Design and Measures section carefully describes the multiple datasets Olzak used in her analyses, her dependent variables (foundings of civilian review boards and the number of officer-involved fatalities), and her independent variables (protests against police brutality and organizational dimensions of protest). Because Olzak's goal is to identify the unique effect of protest activity, she also controls for other variables that may be associated with protest activity in a city—including population size, racial composition, wealth, violent crime rates, and racial disparities in income. The findings section in Figure 16.9 (titled Results) presents analyses of the effect of protests on the founding of civilian review boards and on officer-related fatalities, including the figures presented here and multivariate statistical analyses that adjust for the control variables. As indicated in the combined Discussion and Conclusions section in Figure 16.9, the analysis provides strong evidence that protests influence cities to establish more powerful citizen oversight boards and they lower fatalities in underrepresented communities.

FIGURE 16.9 ■ Sections in a Journal Article

Olzak, Susan. 2021. "Does Protest Against Police Violence Matter? Evidence form U.S. Cities, 1990 through 2019." American Sociological Review 86(6):1066–1099.

INTRODUCTION

Racial tensions have escalated in the wake of deadly encounters with the police, which have spread a climate of fear and distrust in Black and Latino communities and among police officers serving those communities. This study examines whether the current protest movement opposing police brutality has had any tangible effects on two consequences desired by this social movement: implementing civilian review boards (CRBs) and reducing police -involved fatalities. (p. 1067)

THEORETICAL ARGUMENT

My core argument is that protest produces both policy change and changes in police behaviors concerning lethal violence through three key mechanisms…. (p. 1069)

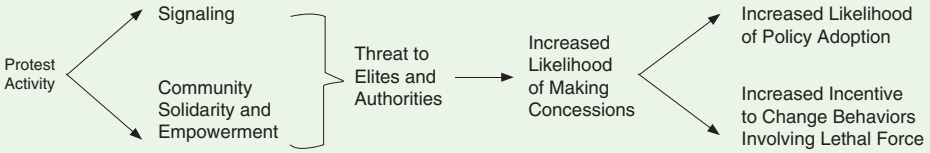

Protest Activity
Signaling
Community Solidarity and Empowerment
Threat to Elites and Authorities
Increased Likelihood of Making Concessions
Increased Likelihood of Policy Adoption
Increased Incentive to Change Behaviors Involving Lethal Force

Figure 1. Mechanisms Connecting Protest Activity to Social Movement Outcomes The argument in Figure 1 implies that protest activates a signaling mechanism and increases community solidarity that empowers disadvantaged groups… raise costs toinaction by elites and provide incentives to make concessions. Such concessions increase the likelihood that policy change will occur, and they provide incentives to change behaviors. (p. 1070)

HYPOTHESES

Hypothesis 1 suggests that cities with more protests against police violence will have a higher likelihood of establishing a CRB, controlling for population composition, levels of violent crime, and racial disparities in income. (p. 1072)

RESEARCH DESIGN AND MEASURES

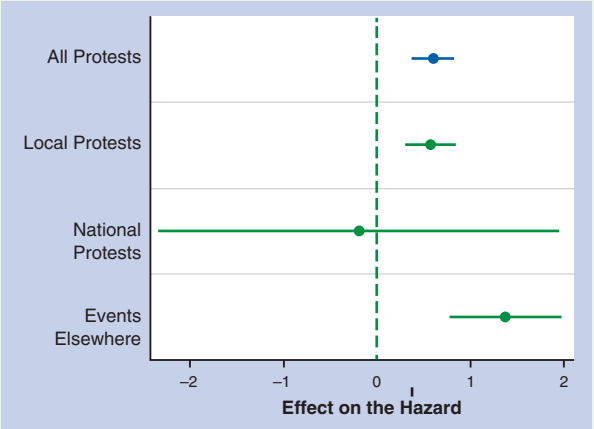
Follow 170 U.S. cities (with populations greater than 100,000) beginning in 1980 through the end of 2018. …I rely on sources of data on protests from Newsbank, which includes news reports from thousands of local newspapers. (p. 1072)

In the first analysis, concessions take the form of establishing a CRB [Community Relations Board]. …The second dependent variable is a count of police-involved fatalities of Black, Latino, and White individuals, 2001 through the end of 2019. (p. 1073)

CRB FOUNDING ANALYSIS

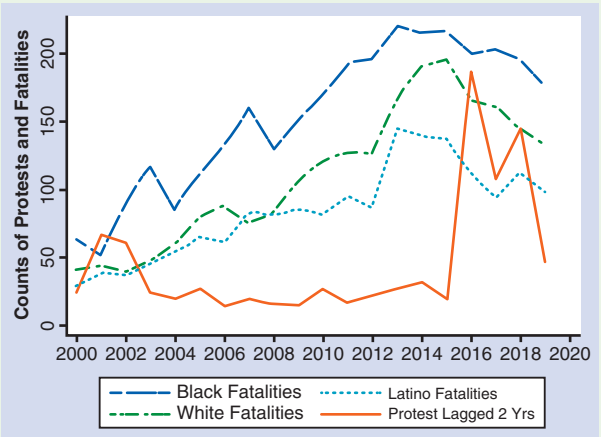
The effect of "national events" is close to zero…. Local protests and protests against police violence that occurred in some other city had strong positive effects on this rate [of founding of community relations boards]. (p. 1079)

OFFICER-RELATED FATALITIES



The results suggest local protest closely tracked the peak years of Black fatalities. (p. 1082)

DISCUSSION AND CONCLUSIONS



My results show that protests influence cities to establish more powerful citizen oversight boards and they lower fatalities in minority communities. (p. 1089)

These basic article sections present research results well, with subsections tailored to the development of the argument and the elements in the analysis. Most journals require a short abstract at the beginning, which summarizes the research question and findings. Most research articles include a general Methodology section that will include subsections on measurement and sampling. A Conclusions section is often used to present the most general conclusions, reflections, and limitations, but most often this is preceded by a general Discussion section; sometimes the two are combined, as in Olzak's article.

Applied Research Reports

Applied research reports are written for a different audience from the professional social scientists and students who read academic journals. Typically, an applied report is written with a wide audience of potential users in mind and to serve multiple purposes. Often, both the

audience and the purpose are established by the agency or other organization that funded the research project on which the report is based. Sometimes, the researcher may use the report to provide a broad descriptive overview of the study findings, which will be presented more succinctly in a subsequent journal article that focuses on a particular research question. An applied report typically provides much more information about a research project than does a journal article and relies primarily on descriptive statistics rather than only those statistics useful for the specific hypothesis tests or exploratory analyses that are likely to be the primary focus of a journal article.

One of the major differences between an applied research report and a journal article is that a journal article must focus on answering a particular research question or questions, whereas an applied report is likely to have the broader purpose of describing a wide range of study findings and attempting to meet the diverse needs of multiple audiences for the research. But a research report that simply describes "findings" without some larger purpose in mind is unlikely to be effective in reaching any audience. Anticipating the needs of the intended audience(s) and identifying the ways in which the report can be useful to them will result in a product that is more likely to be used.

Findings From Living in Boston During COVID-19: Economic Impact

For *Living in Boston During COVID-19*, our National Science Foundation-funded study, our collaborators at the Boston Area Research Initiative (BARI) developed a common template for the seven reports we issued to share project findings with a broad audience. We supplemented these 20-page or so reports with brief "data stories" about important findings that we released on blogs and a video posted on YouTube.

Figure 16.10 displays a portion of the primary project site and introduces the project report on the pandemic's economic impact (report 3), which I will use to illustrate applied research reports.

FIGURE 16.10 ■ COVID in Boston, Homepage ⓡ Boston Area Research Initiative

The **Living in Boston during COVID Survey** was conducted by BARI, the Center for Survey Research at UMass Boston, and the Boston Public Health Commission, with generous support from the National Science Foundation. It is the only survey in the world to capture disparities in the experiences of, attitudes toward, and impacts from the COVID-19 pandemic across the neighborhoods of a city, providing a unique insight into racial and social inequities. We are highlighting the insights from the survey in a series of reports and bite-size data stories.

Reports

Read our first report Living in Boston During COVID: Inequities in Navigating a Pandemic, which highlights how some Bostonians were able to alter their daily activities to minimize their exposure to risk and maximize their health–whereas others were not. Also see our data stories on "Who Had to Go to Work", "Balancing Food Access and Infection Risk", and "Who Rode Transit and Exercised?". You can also watch the video describing the report Living in Boston During COVID on our YouTube page. We also invite you to read this Boston Globe article capturing our first report *Inequities in Navigating a Pandemic*.

Read our second report, Living in Boston During COVID: Fear and Ambivalence, which describes the attitudes and behaviors that might create vulnerabilities for spreading infection in different communities. Also, see our data stories on Attitudes toward Risk & Guidelines: Signs of Ambivalence, High-Risk Behaviors: Acute Neighborhood Differences, Mask Wearing: Self and Others, and Asymptomatic Spread: Fact or Fiction. You can also watch the video describing the report Living in Boston During COVID: Fear and Ambivalence on our YouTube page.

Read our third report, Living in Boston During COVID: Economic Impact, which describes how the pandemic has affected the employment trajectories and economic resources of Boston's residents. Also, see our data stories on Widespread but Variable, Inequities Reinforced and Exacerbated by Pandemic. You can also watch the video describing the report Living in Boston During COVID: Economic Strain our YouTube page.

Source: https://cssh.northeastern.edu/bari/projects/covid-19-in-boston/.

Each report begins with an executive summary, some background information about the project, and then sections on findings related to the report's focus, with tables, charts, and neighborhood diagrams to highlight key results, and then conclusions and a methodological appendix (see Figure 16.11).

FIGURE 16.11 ■ Sections in an Applied Report

Schutt, Russell K., Daniel T. O'Brien, Lee Hargraves, Dan Dooley, Floyd Fowler, Ryan Qi Want, Alina Ristea, Anthony Roman, Mehrnaz Amiri, Sage Gibbons, Hannah Grabowski, and Nikola Kovacevic. 2020. *Living in Boston During COVID-19: Economic Strains*. Boston: Boston Area Research Initiative (BARI), Northeastern University; BEACON Project, Center for Survey Research, University of Massachusetts Boston; Boston Public Health Commission, City of Boston.

EXECUTIVE SUMMARY

In the Summer of 2020, the Boston Area Research Initiative (BARI) at Northeastern University, the Center for Survey Research (CSR) at University of Massachusetts Boston, and the Boston Public Health Commission (BPHC) conducted a survey that captures the experiences of 1626 Bostonians during the first months of the COVID-19 pandemic, including: their ability and tendency to follow social distancing recommendations; attitudes toward regulations; and the economic and personal impacts of the pandemic.

This third report in a series describes how the pandemic has affected the employment trajectories and economic resources of Boston's residents. By examining how these impacts were distributed by race and ethnicity, socioeconomic status, and family composition, we reveal a range of inequities in how different populations have experienced financial disadvantage due to the pandemic.

Main Findings

- **More than four in ten of Boston's residents experienced an adverse economic impact of the pandemic by the time of the survey.**
  - **About one-quarter of those working just before the pandemic spent some time not working after it began**, but seven percent returned to work by the summer.
  - **About one in five of all respondents reported their personal income had declined "a lot,"** while another one in five reported their income had declined a little.
  - Only seven percent reported that their personal income had gone up during the pandemic.

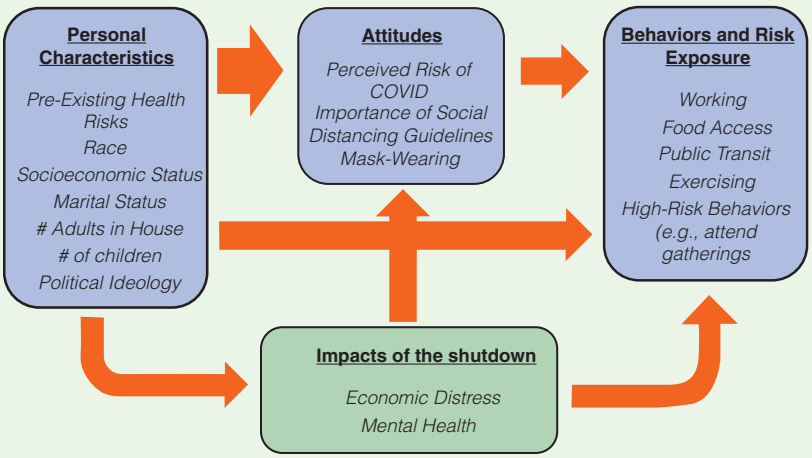
I. INTRODUCTION

1. LIVING IN BOSTON DURING COVID-19: A NEIGHBORHOOD SURVEY

The NSF-Beacon survey captures the experiences of 1626 Bostonians during the first months of the COVID-19 pandemic, including: their ability and tendency to follow social distancing recommendations; attitudes toward regulations; and the economic and personal impacts of the pandemic. It provides unique insights into how these factors varied across the populations and neighborhoods of a single city—something not currently available from any other source, in Boston or otherwise.

The Center for Survey Research (CSR) at University of Massachusetts Boston conducted the survey over the summer, in collaboration with the Boston Area Research Initiative (BARI) at Northeastern University, and the Boston Public Health Commission (BPHC). The National Science Foundation's Human-Environment and Geographical Sciences (HEGS) program provided funding through a grant for rapid-response research (RAPID). The survey used a probability-based random sample stratified by 25 neighborhoods and the results presented here were weighted to match the demographic composition of the city. See Appendix A for more detail on the survey methodology.

This is the third in a series of reports describing key insights from the survey. The series focuses especially on the racial and socioeconomic inequities that have exacerbated—and may continue to exacerbate—differential impacts of the pandemic and the associated shutdown. In doing so, we consider four crucial classes of factors. The first class is personal characteristics, including race, ethnicity, socioeconomic status, pre-existing health, family structure (e.g., number of children), and political ideology. Second are attitudes about the risk of infection and social distancing guidelines, such as mask-wearing. Third are the types of activities that might expose a person to infection. For instance, how often a person goes to work, the grocery store, rides public transit, or visits in other people's houses influences their exposure risk. Fourth, the survey included items on the impacts of the pandemic: employment, economic insecurity, and mental health.



II. TABLE OF ECONOMIC STATUS QUESTIONS

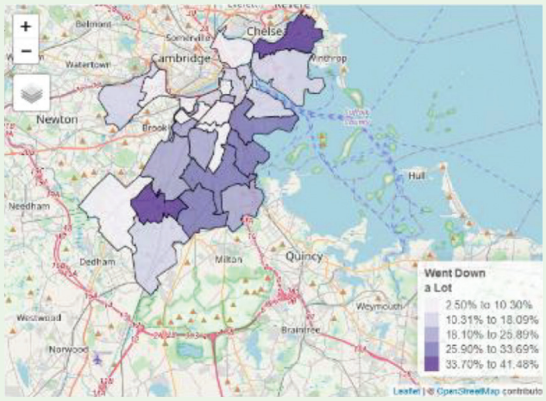
Employment Change
In March 2020 that is before the shutdowns related to the COVID-19 coronavirus began, were you employed at a job or do any kind of work for pay?
In April 2020, were you employed at a job or do any kind of work for pay?
In the last 7 days, were you employed at a job or do any kind of work for pay?
Income Loss: How has your personal income been affected by the COVID-19 coronavirus—has it . . .
Saving: Right Now, do you have enough savings to pay for a $1000 emergency response?
Financial Worries
How worried are you, if at all, with your ability to pay your bills?
How worried are you that you will not be able to make a rent or mortgage payment?
How worried are you that you will run out of food because of a lack of money or resources?

III. FINDINGS ON RACE, ETHNICITY, AND INCOME LOSS

While fewer than one in three white, non-Hispanic respondents reported any income loss as a result of the pandemic, about half of those who identified as Latinx, Black, or Asian did so; those identifying as in some other group reported somewhat more income loss than whites, but not as much as reported by Latinx, Black, and Asian respondents (Figure 9). Latinx and Asian respondents were also least likely to have experienced any income gain as a result of pandemic.

IV. ECONOMIC IMPACT VARIED BY NEIGHBORHOOD

Income loss was also concentrated by neighborhood. In neighborhoods with many affluent residents, like Beacon Hill, Back Bay, the South End, and West Roxbury, few suffered a major loss of income (Figure 12). In some of the less affluent communities like Mattapan and East Boston/Orient Heights, income loss was much more common. But the neighborhood variation only reveals part of the story.



V. CONCLUSIONS

The Living in Boston during COVID-19 survey captures how the economic impact of the pandemic varies across neighborhoods and populations. In this third report we have concentrated on the extent of economic impact and how income loss varied between individuals in relation to their economic vulnerability, race and ethnicity, education, income, and family status. We have also shown how some of these patterns were reflected in differences between neighborhoods. Disparities in economic impact thus paralleled and so compounded existing inequities in economic resources between neighborhoods and social groups. In another report we will examine how the pandemic's disparate economic impact related to differences in exposure to infection risk at work and in the community.

Appendix A: NSF Beacon Survey Methodology

The NSF-Beacon survey is a collaboration of the Boston Area Research Initiative (BARI) at Northeastern University, the Center for Survey Research (CSR) at University of Massachusetts Boston, and the Boston Public Health Commission (BPHC), funded by the National Science Foundation's Human-Environment and Geographical Sciences (HEGS) program through a grant for rapid-response research (RAPID) for collecting ephemeral data during or following a crisis. The survey captures the experiences of 1370 Bostonians during the first months of the COVID-19 pandemic, including ability and tendency to follow social distancing recommendations, attitudes towards regulations, and economic and personal impacts of the pandemic. The design allows for a unique observation of neighborhood-level estimates for these factors.

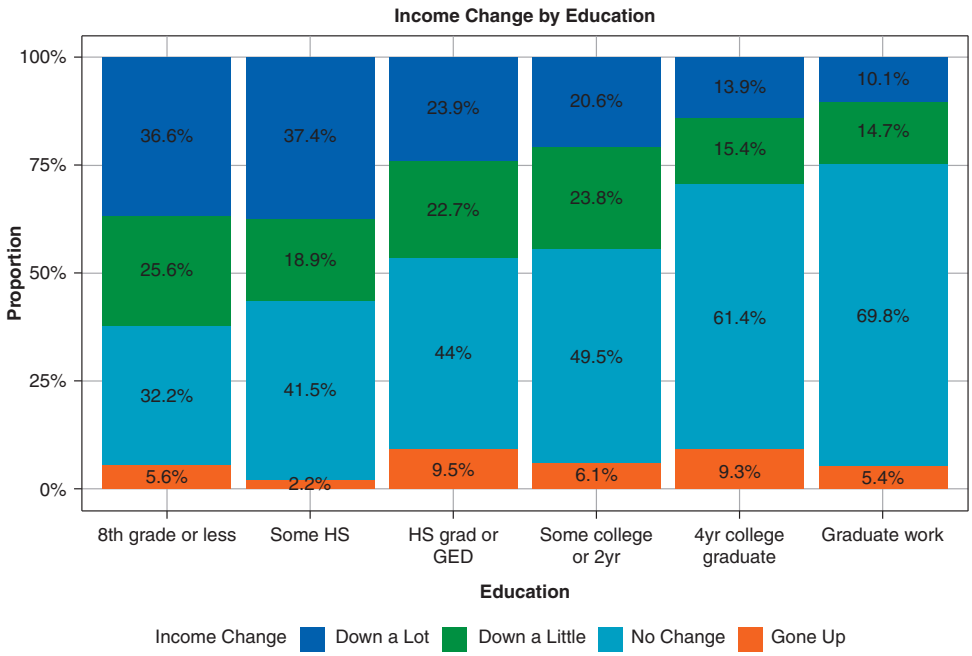
I. Sample Design and Final Sample

The NSF-Beacon survey used a stratified random sample that divided the city of Boston into 25 distinct neighborhoods. The neighborhoods were defined in collaboration with members of the Mayor's Office and other experts based on social, demographic, and historical salience.

Source: Schutt et al. 2020. Living in Boston During COVID-19: Economic Strains. Report #3.

Three separate "data stories" summarized findings about one of the issues addressed in the report. For example, Figure 16.12 includes a selection from the data story, "Economic Impact: Inequities Reinforced."

FIGURE 16.12 ■ Economic Impact: Inequities Reinforced, Living in Boston During COVID19 Data Story



Source: https://cssh.northeastern.edu/bari/economic-impact-inequities-reinforced/.

Framing an Applied Report

What can be termed the **front matter** and the **back matter** of an applied report are important. As in our *Living in Boston During COVID-19* reports, applied reports usually begin with an executive summary: a bulleted list of the study's key characteristics, main findings, and conclusions. Appendixes—the back matter—may present tables containing supporting data or methodological details that were not discussed in the body of the report. Applied research reports also often append a copy of the research instrument(s).

An important principle for the researcher writing for a nonacademic audience is that the findings and conclusions should be engaging and clear. You can see how I did this in a complete report from a class research project I designed with my graduate methods students (and in collaboration with several faculty knowledgeable about substance abuse) (see Figure 16.13). These report excerpts indicate how I summarized key findings in an executive summary (Schutt et al. 1996:iv), emphasized the importance of the research in the introduction (Schutt et al. 1996:1), used formatting and graphing to draw attention to particular findings in the body of the text (Schutt et al. 1996:5), and tailored recommendations to our own university context (Schutt et al. 1996:26).

Research Videos

The popularity of uploading videos to YouTube and other platforms indicates how appealing it is to convey information through audiovisual means. A video with an overview of a research project or a summary of a report may make the research more widely accessible. Figure 16.14

FIGURE 16.13 ■ Student Substance Abuse, Report Excerpts

EXECUTIVE SUMMARY

- Rates of substance abuse were somewhat lower at UMass–Boston than among nationally selected samples of college students.
- Two-thirds of the respondents reported at least one close family member whose drinking or drug use had ever been of concern to them—one-third reported a high level of concern.
- Most students perceived substantial risk of harm due to illicit drug use, but just one-quarter thought alcohol use posed a great risk of harm.

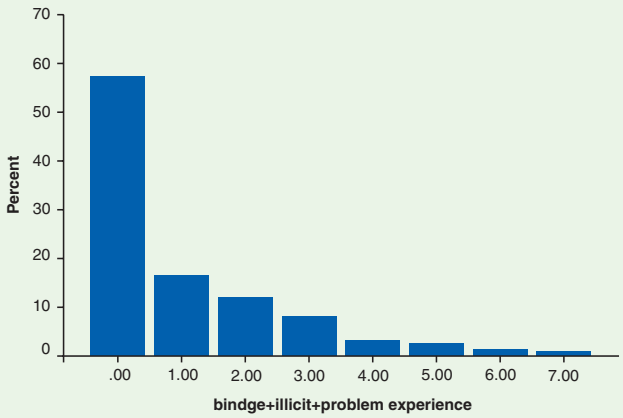
INTRODUCTION

Binge drinking, other forms of alcohol abuse, and illicit drug use create numerous problems on college campuses. Deaths from binge drinking are too common and substance abuse is a factor in as many as two-thirds of on-campus sexual assaults (Finn 1997; National Institute of Alcohol Abuse and Alcoholism 1995). College presidents now rate alcohol abuse as the number one campus problem (Wechsler, Davenport, Dowdall, Moeykens, & Castillo 1994) and many schools have been devising new substance abuse prevention policies and programs. However, in spite of increasing recognition of and knowledge about substance abuse problems at colleges as a whole, little attention has been focused on substance abuse at commuter schools.

FINDINGS

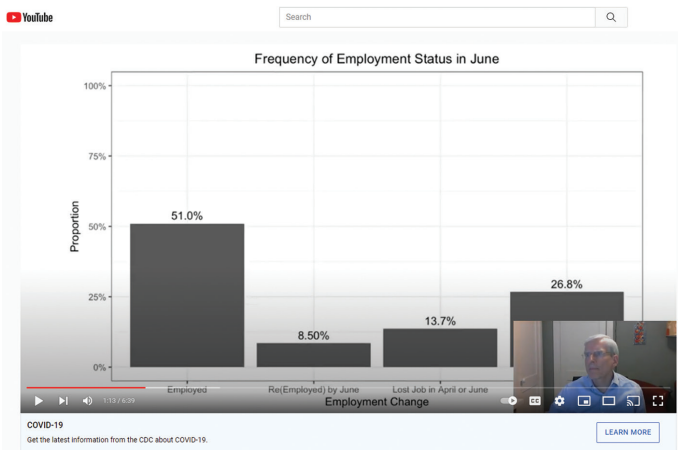
The composite index identifies 27% of respondents as at risk of substance abuse (an index score of 2 or higher). One-quarter reported having smoked or used smokeless tobacco in the past two weeks.

27% of respondents were identified as at risk of substance abuse.



Source: Schutt et al. (1996).

FIGURE 16.14 ■ Screenshot From a Research Video



presents a screenshot from a research video in which I highlighted some of the findings from the COVID-19 report. You can judge for yourself how helpful it is by watching the video at the following link: https://www.youtube.com/watch?v=VKFRRgeBWwA.

Research Posters and Presentations

Research posters have long been a popular tool for summarizing research projects and findings at professional meetings in medicine and psychology, and professional associations in sociology now often include one or more poster sessions. A well-designed poster can provide much of the key information that would otherwise appear in a paper or presentation, while the poster session format can allow meeting participants to question presenters at length about their work and perhaps exchange information about their own or just consider the findings from a distance.

SAGE author Vera Toepoel (2016:217) provides helpful guidelines:

- Completely fill the poster, avoiding clutter or overwhelming appearance.
- Include no more than 15 lines per page.
- Use bullets, but brief, on one point only, and not too many in one list.
- Use a font large enough for reading 3–6 feet away.
- Graphs and visuals usually are better than plain text.
- Use software designed to create posters.
- Ensure a customary flow, using top left to bottom right and numbered sections.
- Place most important or eye-catching information (usually results) in the center.

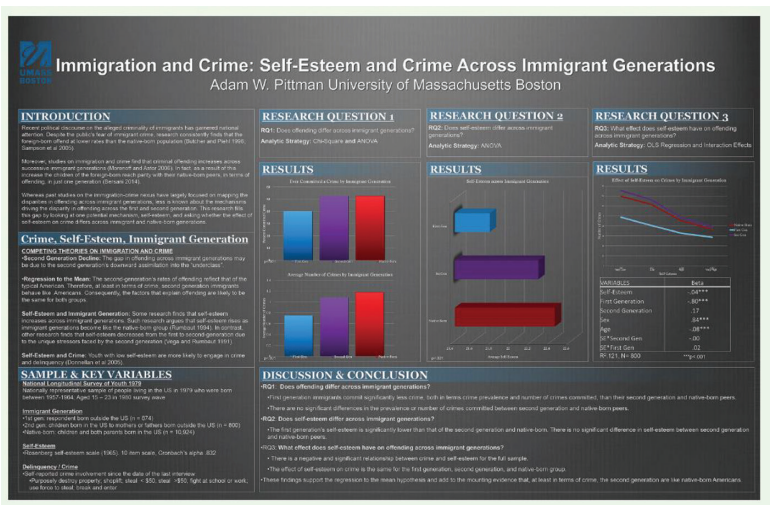
Most of the same guidelines apply for creating slides to use in a presentation, but you can use many more slides than can fit within a single poster. One slide per minute of presentation is a good rule of thumb. Avoid lots of text in a slide, using short bulleted lists to highlight key points in text and charts or other images to convey data rather than numerical tables when possible. Don't write out everything you plan to say on the slides, and avoid reading the slides verbatim. Practicing talking through your slides before your presentation so you feel comfortable making the key points on each slide and making transitional comments as you advance through the slides.

The University of Texas provides guidelines on their website (https://ugs.utexas.edu/our/poster/samples) and several good examples. Dr. Adam Pittman's poster (see Figure 16.15) on immigration and crime is a good model. You can create a poster like this just using PowerPoint (with a large printer), but it's much easier if you use software designed for this purpose.

Presentations using PowerPoint or other presentation programs have become the norm for displaying research findings in the classroom, in professional meetings, and in public. When you display your research in this way, it's important to consider how to convey the key information without overloading or confusing your audience. Once again, SAGE author Vera Toepoel (2016:214) gives some straightforward guidelines:

- Make sure a slide contains only one idea, theory, or example.
- Do not spend too much time on each slide (1–2 minutes maximum).
- Give each slide a title that clearly explains what you are talking about.

FIGURE 16.15 ■ Poster Example

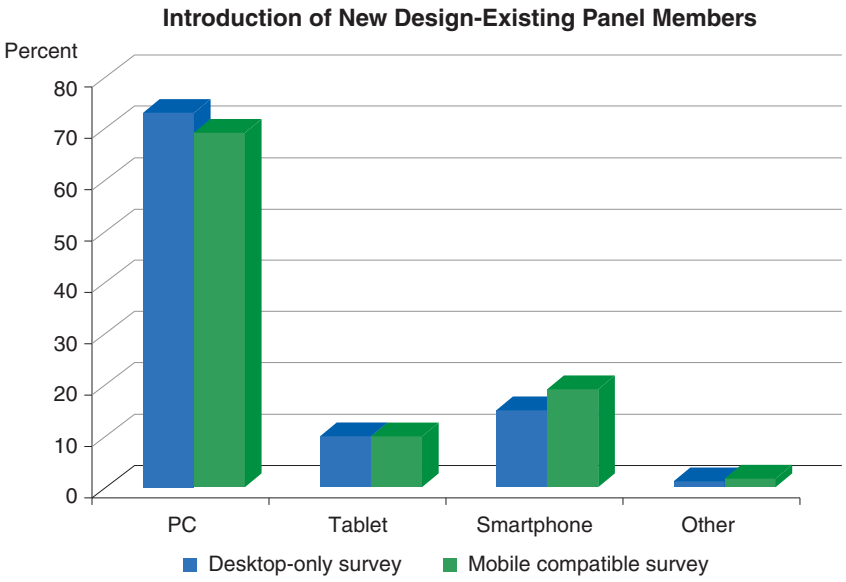


Source: Adam W. Pittman, University of Massachusetts, Boston.

- Do not use too many words on one slide (7 words × 7 rows maximum).
- Use bullets to begin each separate point.
- Use cartoons, animation, fonts, and colors to increase the clarity of the point without impeding readability or looking messy.
- Cite sources and include references on the last slide.

Figure 16.16 gives an example of a slide from one of Toepoel's own PowerPoint presentations with two colleagues.

FIGURE 16.16 ■ Attractive Chart in PowerPoint



Source: Reprinted with permission from Vera Toepoel.

Reporting Quantitative and Qualitative Research

The requirements for good research reports are similar in many respects for quantitative and qualitative research projects. Every research report should include good writing, a clear statement of the research question, an integrated literature review, and the presentation of key findings with related discussion, conclusions, and limitations. A research report of a mixed-methods study can integrate the results of analyses of both types of data. However, the differences between qualitative and quantitative research approaches mean that it is often desirable for research reports based on qualitative research to diverge in some respects from those reflecting quantitative research.

Reports based on qualitative research should be enriched in each section with elements that reflect the more holistic and reflexive approach of qualitative projects. The introduction should include background about the development of the researcher's interest in the topic, while the literature review should include some attention to the types of particular qualitative methods used in prior research. The methodology section should describe how the researcher gained access to the setting or individuals studied and the approach used for managing relations with research participants. The presentation of findings in qualitative studies may be organized into sections reflecting different themes identified in interviews or observational sessions. Quotes from participants or from observational notes should be selected to illustrate these themes, although qualitative research reports differ in the extent to which the researcher presents findings in summary form or uses direct quotes to identify key issues. The findings sections in a qualitative report may alternate between presentations of quotes or observations about the research participants, the researcher's interpretations of this material, and some commentary on how the researcher reacted in the setting, although some qualitative researchers will limit their discussion of their reactions to the discussion section.

Reports on mixed-methods projects should include subsections in the methods section that introduce each method and then distinguish findings from qualitative and quantitative analyses in the findings section. Some mixed-methods research reports may present analyses that use both qualitative and quantitative data in yet another subsection, but others may just discuss implications of analyses of each type for the overall conclusions in the discussions and conclusions sections (Dahlberg, Wittink, and Gallo 2010:785–791). When findings based on each method are presented, it is important to consider explicitly both the ways in which the specific methods influenced findings obtained with those methods and to discuss the implications of findings obtained using both methods for the overall study conclusions.

RESEARCH IN THE NEWS
WHO HAS ENOUGH CASH TO GET THROUGH THE CORONAVIRUS CRISIS?

Many Americans lived paycheck to paycheck even before the pandemic, struggling to afford housing and childcare costs, student loan payments, medical costs, and more. But as you know, the pandemic disrupted work opportunities for many. Only 47% of Americans felt they had enough to cover just three months of expenses, and of course, the pandemic prevented many from working for much longer than that. For lower-income Americans, expenses far exceed income. Some found themselves having trouble concentrating and sleeping due to financial worries.

For Further Thought?

1. What type of study would you recommend for investigating the impact of the pandemic on economic security? What would have been some of the challenges to implementing your study design?

2. In what ways could you report the results of research on the value of government financial support among lower-income Americans in order to inform the public? Do you think that researchers should try to influence the public debate through their reports or just stick to publishing in academic journals?

News source: Quart, Alissa, and Yaryna Serkez. 2020. "Who Has Enough Cash to Get Through the Coronavirus Crisis." *The New York Times*, April 26:2.

ETHICS, POLITICS, AND RESEARCH REPORTS

The researcher's ethical duty to be honest and open becomes paramount when reporting research results. Here are some guidelines:

- *Provide an honest accounting of how the research was carried out and where the initial research design had to be changed.* Readers do not have to know about every change you made in your plans and each new idea you had, but they should be informed about major changes in hypotheses or research design. If important hypotheses were not supported, acknowledge this rather than conveniently forget to mention them (Brown and Hedges 2009:383). If a different approach to collecting or analyzing the data could have led to different conclusions, this should be acknowledged in the limitations section (Bergman 2008:588–590).
- *Evaluate honestly the strengths and weaknesses of your research design.* Systematic evaluations suggest that the stronger the research design from the standpoint of establishing internal (causal) validity, the weaker the empirical support that is likely to be found for hypothesized effects (compare Weisburd, Lum, and Petrosino 2001:64). Thus, it is less likely that a hypothesis tested with a stronger randomized experimental design will be supported than when the same hypothesis is tested based on correlations between variables measured in a cross-sectional survey.
- *Refer to prior research and interpret your findings within the body of literature resulting from that prior research.* Your results are likely to be only the latest research conducted to investigate a research question that others have studied. It borders on unethical practice to present your findings as if they are the only empirical information with which to answer your research question, yet many researchers commit this fundamental mistake (Bergman 2008:599). For example, a systematic evaluation of citation frequency in articles reporting clinical trial results in medical journals found that, on average, just 21% of the available prior research was cited (for trials with at least three prior articles that could have been cited) (Robinson and Goodman 2011:53). The result of such omission is that readers may have no idea whether your own research supports a larger body of evidence or differs from it—and so should be subject to even greater scrutiny.
- *Maintain a full record of the research project so that questions can be answered if they arise.* Many details will have to be omitted from all but the most comprehensive reports, but these omissions should not make it impossible to track down answers to specific questions about research procedures that may arise during data analysis or presentation. Tests of relationships that were conducted but not included in the report should be acknowledged.
- *Avoid "lying with statistics" or using graphs to mislead.* (See Chapter 9 for more on this topic.) There is a more subtle problem to be avoided, which is "cherry-picking" results

to present. Although some studies are designed to test only one hypothesis involving variables that each are measured in only one way, many studies collect data that can be used to test many hypotheses, often with alternative measures. If many possible relationships have been examined with the data collected and only those found to yield a statistically significant result are reported, the odds of capitalizing on chance findings are multiplied. This is a major temptation in research practice and has the unfortunate result that most published findings are not replicated or do not stand up to repeated tests over time (Lehrer 2010:57). Every statistical test presented can be adequately understood only in light of the entire body of statistical analyses that led to that particular result.

- *Acknowledge the sponsors of the research.* This is important partly so that others can consider whether this sponsorship may have tempted you to bias your results in some way. Whether you conducted your research for a sponsor or together with members of an underserved community, give research participants an opportunity to comment on your main findings before you release them to the public. Consider revising your report based on their suggestions or, if you disagree with their suggestions, acknowledging their comments in footnotes at relevant points in your report or as an appendix to it (Bledsoe and Hopson 2009:392).
- *Thank staff who made major contributions.* This is an ethical as well as a political necessity. Let's maintain our social relations!
- *Be sure that the order of authorship for coauthored reports is discussed in advance and reflects agreed-on principles.* Be sensitive to coauthors' needs and concerns and to the extent of each coauthor's contribution.

Ethical research reporting should not mean ineffective reporting. You need to tell a coherent story in the report and avoid losing track of the story in a thicket of minuscule details. You do not need to report every twist and turn in the conceptualization of the research problem or the conduct of the research. But be suspicious of reports that don't seem to admit to the possibility of any room for improvement. Social science is an ongoing enterprise in which one research report makes its most valuable contribution by laying the groundwork for another, more sophisticated, research project. Highlight important findings in the research report but also use the research report to point out what are likely to be the most productive directions for future researchers.

Ethical research reporting should mean reporting findings regardless of whether they support your hypotheses, but that is often not the case. When you "don't find anything" or your hypotheses are not supported, this can be important information for other researchers, but it will be very hard—even impossible—to have an article about the "null findings" accepted by a journal. Journals (and authors) like to publish articles that show something new about social phenomena (Gough, Oliver, and Thomas 2017:99). There can also be a concern that poor decisions in the research methods used may have led to the lack of anticipated effects. Over time, this tendency means that the peer-reviewed journal literature can have a **publication bias** toward research that shows hypothesized effects were found rather than showing that the anticipated effects were not found. Since hypotheses are often not supported, researchers joke about wishing there was a *Journal of Null Findings* in which they could publish these undesired outcomes. Concerns about publication bias in medical research led the U.S. Congress in 1997 to pass a law requiring the registration of clinical trials (health-related experiments) before enrolling research subjects (U.S. National Library of Medicine n.d.). Many medically oriented journals now require

that all clinical trials reported in an article submitted for possible publication were registered in the publicly available database maintained by the U.S. National Library of Medicine at clinicaltrials. gov. Authors must confirm at the time they submit an article to one of these journals that their research project was registered in the database.

Even following appropriate guidelines such as these, however, will not prevent controversy and conflict over research on sensitive issues. The late sociologist Peter Rossi (1999) recounted the controversy that arose when he released a summary of findings obtained in his 1989 study of homeless persons in Chicago (seeChapter 5). Despite important findings about the causes and effects of homelessness, media attention focused on the fact that Rossi's summary included a markedly smaller estimate of the numbers of homeless persons in Chicago compared with the "guesstimates" that had been publicized by local advocacy groups. "Moral of the story: Controversy is news, to which careful empirical findings cannot be compared" (Rossi 1999:2).

Does this mean that ethical researchers should avoid political controversy by sidestepping media outlets for their work? Many social scientists argue that the media offers one of the best ways to communicate the practical application of sociological knowledge, so when we avoid these opportunities, "some of the best sociological insights never reach policy makers because sociologists seldom take advantage of useful mechanisms to get their ideas out" (Wilson 1998:435).

The sociologist William Julius Wilson (1998:438) urges the following principles for engaging the public through the media:

1. Focus on issues of national concern, issues that are high on the public agenda.
2. Develop creative and thoughtful arguments that are clearly presented and devoid of technical language.
3. Present the big picture whereby the arguments are organized and presented so that the readers can see how the various parts are interrelated.

Ultimately, each researcher or research team must make a decision about the most appropriate and important outlets for their work.

PLAGIARISM

It may seem depressing to end a book on research methods with a section on **plagiarism**, but it would be irresponsible to avoid the topic. Of course, you may have a course syllabus detailing instructor or university policies about plagiarism and specifying the penalties for violating that policy, so I'm not simply going to repeat that kind of warning. You probably realize the practice of selling term papers is revoltingly widespread (my search of "term papers for sale" on Google returned 369,000 websites on April 21, 2022)—for at least the first 8 pages of "hits," every site provided an opportunity to buy a paper, not a criticism of the practice. Instead of repeating the warnings, I will use this section to review the concept of plagiarism and to show how that problem connects to the larger issue of the integrity of social research. When you understand the dimensions of the problem and the way it affects research, you should be better able to detect plagiarism in other work and avoid it in your own.

You learned in Chapter 3 that maintaining professional integrity—honesty and openness in research procedures and results—is the foundation for ethical research practice. When it comes to research publications and reports, being honest and open means above all else avoiding plagiarism—that is, presenting as one's own the ideas or words of another person or persons for academic evaluation without proper acknowledgment (Hard, Conway, and Moran 2006:1059).

An increasing body of research suggests that plagiarism is a growing problem on college campuses. Jason Stephen et al. (2007:243) found in a web survey of self-selected students at two universities that one quarter acknowledged having plagiarized a few sentences (24.7%) or a complete paper (0.3%) in coursework within the past year (many others admitted to other forms of academic dishonesty, such as copying homework). Hard et al. (2006) conducted an anonymous survey in selected classes in one university, with almost all students participating, and found much higher plagiarism rates: 60.6% reported that they had copied "sentences, phrases, paragraphs, tables, figures, or data directly or in slightly modified form from a book, article, or other academic source without using quotation marks or giving proper acknowledgment to the original author or source" (p. 1069), and 39.4% reported that they had "copied information from Internet websites and submitted it as [their] work" (p. 1069).

So the plagiarism problem is not just about purchasing term papers—although that is really about as bad as it gets (Broskoske 2005:1); plagiarism is also about what you do with the information you obtain from a literature review or inspection of research reports. And rest assured this is not only about student papers; it also is about the work of established scholars and social researchers who publish reports you want to rely on for accurate information. Several noted researchers have been accused of plagiarizing passages they used in popular books or academic articles; some have admitted to not checking the work of their research assistants, to not keeping track of their sources, or to being unable to retrieve the data they claimed they had analyzed. Serious plagiarism by academic researchers has been identified from New York to Witwatersrand (South Africa), Birmingham (England), and Beijing (Israel 2014:158–159). Whether the cause is cutting corners to meet deadlines or consciously fudging facts, the effect is to undermine the trustworthiness of social research.

What about Wikipedia? With over 5 million Wikipedia articles available in English in 2017 (and 40 million in a total of 293 languages), you can find an article on just about any topic of interest. Of course, it is nothing but plagiarism to copy text out of a Wikipedia article and paste it into your paper without proper attribution. But what about using Wikipedia articles as sources for your literature review? Research published in 2005 in the prestigious journal *Nature* found that Wikipedia articles were almost as accurate as those in *Encyclopedia Britannica*—which uses experts to author its articles (Giles 2005). The sheer number of people involved in Wikipedia—contributing text and editing it—generates a self-correcting process that works pretty well, even without expert control (Abernathy 2017:42). Yet it's still not appropriate to use Wikipedia or any other encyclopedia as a source in an academic paper or article. The problem is their articles are distillations of findings from the original sources, and you have no way of knowing how accurately they represent the original findings as they apply to your own research question. That's the hard work you have to do yourself.

Now that you are completing this course in research methods, it's time to think about how to do your part to reduce the prevalence of plagiarism and ensure that you report accurately the findings of any research you conduct. Of course, the first step is to maintain careful procedures for documenting the sources that you rely on for your own research and papers, but you should also think about how best to reduce temptations among others. After all, what people believe about what others do is a strong influence on their own behavior (Hard et al. 2006:1058).

Reviewing the definition of plagiarism and how your discipline's professional association enforces its policies against the practice is an important first step. These definitions and procedures reflect a collective effort to help social scientists maintain standards throughout the discipline. Awareness is the first step (American Sociological Association 1999:19).

Sociologists have an obligation to be familiar with their Code of Ethics, other applicable ethics codes, and their application to sociologists' work. Lack of awareness or misunderstanding of an ethical standard is not, in itself, a defense to a charge of unethical conduct.

The American Sociological Association (ASA)'s (1999) *Code of Ethics* includes an explicit prohibition of plagiarism:

> 14. Plagiarism
>
> (a) In publications, presentations, teaching, practice, and service, sociologists explicitly identify, credit, and reference the author when they take data or material verbatim from another person's written work, whether it is published, unpublished, or electronically available.
>
> (b) In their publications, presentations, teaching, practice, and service, sociologists provide acknowledgment of and reference to the use of others' work, even if the work is not quoted verbatim or paraphrased, and they do not present others' work as their own whether it is published, unpublished, or electronically available. (p. 16)

The next step toward combating the problem and temptation of plagiarism is to keep focused on the goal of social research methods: investigating the social world. If researchers are motivated by a desire to learn about social relations, to understand how people understand society, and to discover why conflicts arise and how they can be prevented, they will be as concerned with the integrity of their research methods as are those who read and use the results of their research. Throughout *Investigating the Social World*, you have been learning how to use research processes and practices that yield valid findings and trustworthy conclusions. Failing to report honestly and openly on the methods used or sources consulted derails progress toward that goal.

It works the same as with cheating in school. When students are motivated only by the desire to "ace" their tests and receive better grades than others, they are more likely to plagiarize and use other illicit means to achieve that goal. Students who seek first to improve their understanding of the subject matter and to engage in the process of learning are less likely to plagiarize sources or cheat on exams (Kohn 2008:6–7). They are also building the foundation for becoming successful researchers who help others understand our social world.

CONCLUSIONS

A well-written research article or report requires (to be just a bit melodramatic) blood, sweat, and tears and more time than you will, at first, anticipate. But the process of writing one will help you write the next. And the issues you consider, if you approach your writing critically, will be sure to improve your subsequent research projects and sharpen your evaluations of other investigators' research projects.

Good critical skills are essential when evaluating research reports, whether your own or those produced by others. There are *always* weak points in any research, even published research. It is an indication of strength, not weakness, to recognize areas where one's own research needs to be or could have been improved. And it is not a question of sharpening your knives and going for the jugular. You need to be able to weigh the strengths and weaknesses of every research result and to evaluate a study for its contribution to understanding the social world—not whether it gives a definitive answer for all time.

But this is not to say that anything goes. Much research lacks one or more of the three legs of validity—measurement validity, causal validity, or generalizability—and some contributes more confusion than understanding about the social world. Top journals generally maintain very high standards, partly because they have good critics in the review process and

distinguished editors who make the final acceptance decisions. But some daily newspapers do a poor job of screening, and research reporting standards in many popular magazines, TV shows, and books are often abysmally poor. Keep your standards high and your views critical when reading research reports, but not so high or so critical that you turn away from studies that make tangible contributions to understanding the social world—even if they don't provide definitive answers. And don't be so intimidated by the need to maintain high standards that you shrink from taking advantage of opportunities to conduct research yourself.

The growth of social science methods from their infancy to adolescence, perhaps to young adulthood, ranks as a key intellectual accomplishment of the 20th century. Opinions about the causes and consequences of homelessness no longer need to depend on the scattered impressions of individuals; criminal justice policies can be shaped by systematic evidence of their effectiveness; and changes in the distribution of poverty and wealth in populations can be identified and charted. Employee productivity, neighborhood cohesion, and societal conflict may each be linked to social structures, psychological processes, and economic strains.

But we have seen in recent years that the progress of social science and of science in general is not assured. The modern information environment, including the web, social media, and smartphones, has made an incredible wealth of information available to almost everyone, but it has also generated extraordinary means for disseminating disinformation and supporting foolish ideas about the social world that have no grounding in scientific methods whatsoever. Anyone who seeks to encourage others to think like they do has readily available tools for trying to convince others that they are "right"—without any regard for the methods used to generate the information or the validity of their conclusions. After all, learning how to investigate the social world takes time and effort!

But it's also on us. Social research methods are no more useful than the commitment of researchers to their proper application. Research methods, like all knowledge, can be used poorly or well, for good purposes or bad, when appropriate or not. A claim that a belief is based on social science research itself provides no extra credibility. As you have learned throughout this book, we must first learn which methods were used, how they were applied, and whether interpretations square with the evidence. To investigate the social world, we must keep in mind the lessons of research methods. Doing so will help build a better social world in the 21st century.

KEY TERMS

Back matter (p. 575)
Data visualization (p. 563)
Effect size (p. 566)
Front matter (p. 575)
Meta-analysis (p. 565)
Plagiarism (p. 584)
Publication bias (p. 583)
Reverse outlining (p. 561)

HIGHLIGHTS

LO 16.1 Critique your own approach to writing papers and reports in relation to the principles formulated by experienced writers.

- The goal of research is not just to discover something but also to communicate that discovery to a larger audience: other social scientists, government officials, your teachers, the general public—perhaps several of these audiences.

- Consider the following principles formulated by experienced writers (Booth, Colomb, and Williams 1995:150–151):
  - Respect the complexity of the task and don't expect to write a polished draft in a linear fashion. Your thinking will develop as you write, causing you to reorganize and rewrite.
  - Leave enough time for dead ends, restarts, revisions, and so on, and accept the fact that you will discard much of what you write.
  - Write as fast as you comfortably can. Don't worry about spelling, grammar, and so on until you are polishing things up.
  - Ask anyone whom you trust for reactions to what you have written.
  - Write as you go along, so you have notes and report segments drafted even before you focus on writing the report.
- Reverse outlining: Outlining the sections in an already written draft of a paper or report to improve its organization in the next draft.
- All write-ups of research should be revised several times and critiqued by others before being presented in final form. Reverse outlining can help in this process.

LO 16.2 Illustrate several types of charts used to display research.

- Chart Builder is an option in SPSS that be used to create figures displaying the association between two or three variables.
- Understanding is most likely to be achieved in a data display when it is
  - Trustworthy: The basic pattern is not distorted to change the viewer's impression.
  - Accessible: It should be easy for the viewer to see the key elements in the data and the relationships between them. Decisions about how simple a display needs to be have to take into account the expertise of the intended audience and the time they will have to view the display.
  - Elegant: The design should attract and sustain attention. It's easy to overdo it, but try to make the design engaging as well as informative.

LO 16.3 Argue for using meta-analysis as a tool for advancing science.

- Meta-analysis is a quantitative method for identifying patterns in findings across multiple studies of the same research question.
- Meta-analysis is not appropriate for evaluating results from qualitative studies or from multiple studies that used different methods or measured different dependent variables.
- Eligibility criteria must be specified carefully to determine which studies to include and which to omit as too different.
- The effect size statistic is the key to capturing the association between the independent and dependent variables across multiple studies.
- The meta-analytic approach to synthesizing research findings can result in much more generalizable findings than those obtained with just one study.
- Before we can place any confidence in the results of a meta-analysis, we must be confident that all (or almost all) relevant studies were included, that the information we need to analyze was included in all (or most) of the studies, and that the studies do not have major methodological weaknesses.

LO 16.4 Compare the advantages of reporting research through journal articles, applied research reports, research videos, and posters.

- Research articles should be evaluated systematically, using the review guide in Appendix A and considering the interrelations between the design elements.
- Journal articles must pass a peer review by other social scientists and are often much improved in the process.
- Applied research reports are constrained by the expectations of the research sponsor; an advisory committee from the applied setting can help avoid problems.
- Typically, an applied report is written with a wide audience of potential users in mind and serves multiple purposes.
- Applied reports begin with an Executive Summary that presents the highlights from each section of the report, including major findings, in a brief format, often using bullets. The front and back matter are also important, including the table of contents and a methodological appendix.
- One of the major differences between an applied research report and a journal article is that a journal article must focus on answering a particular research question, whereas an applied report is likely to have the broader purpose of describing a wide range of study findings and attempting to meet the diverse needs of multiple audiences for the research.
- Research posters have long been a popular tool for summarizing research projects and findings at professional meetings in medicine and psychology, and professional associations in sociology now often include one or more poster sessions.
- Presentations using PowerPoint or other presentation programs have become the norm for displaying research findings in the classroom, in professional meetings, and in public.
  - Make sure that a slide only contains one idea, theory, or example.
  - Do not spend too much time on each slide (1–2 min. maximum).
  - Give your slides a title that clearly explains what you are talking about.
  - Do not use too many words on one slide (7 words × 7 rows max).
  - Use bullet points for each separate point.
  - Use cartoons, animation, fonts, and colors to increase the clarity of the point without impeding readability or looking messy.
  - Cite sources and include references on the last slide.
- A video on YouTube or another platform with an overview of a research project or a summary of a report may make the research more widely accessible.

LO 16.5 Discuss examples of ethical guidelines being followed in an academic journal article and of failures to follow ethical guidelines.

- Provide an honest accounting of how the research was carried out and where the initial research design had to be changed.
- Evaluate honestly the strengths and weaknesses of your research design.
- Refer to prior research and interpret your findings within the body of literature resulting from that prior research.
- Maintain a full record of the research project so that questions can be answered if they arise.

- Avoid "lying with statistics" or using graphs to mislead.
- Acknowledge the sponsors of the research.
- Thank staff who made major contributions.
- Be sure that the order of authorship for coauthored reports is discussed in advance and reflects agreed-on principles.
- Publication bias results from the greater likelihood of acceptance of articles reporting confirmation of hypothesized effects.

LO 16.6 Appraise the motivations for plagiarism and the ways of minimizing them.

- Plagiarism is a grievous violation of scholarly ethics. All direct quotes or paraphrased material from another author's work must be appropriately cited.
- Social scientists are obligated to evaluate the credibility of information obtained from any source before using it in their research reports.
- Plagiarism: Presenting as one's own the ideas or words of another person or persons for academic evaluation without proper acknowledgment.
- An increasing body of research suggests that plagiarism is a growing problem on college campuses.
- The first step in avoiding plagiarism is to maintain careful procedures for documenting the sources that you rely on for your own research and papers, but you should also think about how best to reduce temptations among others.
- The next step toward combating the problem and temptation of plagiarism is to keep focused on the goal of social research methods: investigating the social world.

DISCUSSION QUESTIONS

1. A good place to start developing your critical skills would be with reading Susan Olzak's (2021) article in the *American Sociological Review* that I discussed in this chapter. As you read it, answer the article review questions in Appendix A. How effective is the article in conveying the design and findings of the research? Could the article's organization be improved at all? Are there bases for disagreement about the interpretation of the findings? Did reading the full article increase your opinion of its value?

2. Compare and contrast the goals and features of academic journal articles and applied research reports. What trade-offs are required to present research results in a form that can be understood by those without academic training? What steps can be taken to reduce the chance that results are not misrepresented when details are omitted from an applied research report?

3. How well have you followed guidelines for presenting research slides in your own presentations? Which guidelines have you seen others violate?

4. Prior research on a research question may be reviewed in a traditional narrative research review, in a systematic literature review, and/or in a meta-analysis. When do you feel each of these three approaches is most appropriate? Explain your reasoning.

5. Academic journals tend not to publish articles that test a hypothesis and find no support for it. Why is this considered to be publication bias, and what would you recommend doing about it, if anything?

SPSS EXERCISES

1. What personal characteristics help to explain whether people are more supportive of strengthening "law and order" or more in favor of reducing bias in the criminal justice system? First request the frequencies for each of the variables to be included in the analysis. Then generate the crosstabs between STRVBIAS_2 and the rest of these variables: AGE7, FEMALE, REGION4, RACED, RELIG2020, EDUC6, and INCOME4 (specify SRVBIAS_2 as the column variable and request cell counts and column percents).
2. Present your frequencies tables from #1 using approaches you learned about in Chapter9 or Chapter 16. Use at least some charts, rather than simply tables with numerical results.
3. Now present your crosstabulation tables using approaches you learned about in Chapter 9 or Chapter 16. Use at least some charts, rather than simply tables with numerical results.
4. Write a short report based on these analyses, including the data displays you have just prepared. Include in your report a brief introduction of the research question. In a short methods section, I describe the basic methods used in the GSS 2020 (NORC 2020), and list the questions you used in the analysis. Provide a short summary of each table and chart. Add an integrated discussion and interpretation of what you have found. In a conclusions section, identify limitations and include some suggestions for additional research on sources of attitudes toward criminal justice policies.

APPENDIX A: QUESTIONS TO ASK ABOUT A RESEARCH ARTICLE*

1. What is the basic research question or problem? Try to state it in just one sentence. (Chapter 2)
2. Is the purpose of the study descriptive, exploratory, explanatory, or evaluative? Did the study have more than one purpose? (Chapter 1)
3. Was a theoretical framework presented? What was it? Did it seem appropriate for the research question addressed? Can you think of a different theoretical perspective that might have been used? What philosophy guides the research? Is this philosophy appropriate to the research question? (Chapters 1, 2)
4. What prior literature was reviewed? Was it relevant to the research problem? Was it relevant to the theoretical framework? Does the literature review appear to be adequate? Are you aware of (or can you locate) any important studies that have been omitted? (Chapter 2)
5. What features identified the study as deductive or inductive? Do you need additional information in any areas to evaluate the study? (Chapters 1, 2)
6. Did the study seem consistent with current ethical standards? Were any trade-offs made between different ethical guidelines? Was an appropriate balance struck between adherence to ethical standards and use of the most rigorous scientific practices? (Chapter 3)
7. Were any hypotheses stated? Were these hypotheses justified adequately in terms of the theoretical framework and in terms of prior research? (Chapter 2)
8. What were the independent and dependent variables in the hypothesis or hypotheses? Did these variables reflect the theoretical concepts as intended? What direction of association was hypothesized? Were any other variables identified as potentially important? (Chapter 2)
9. What were the major concepts in the research? How and how clearly were they defined? Were some concepts treated as unidimensional that you think might best be thought of as multidimensional? (Chapter 4)
10. Did the instruments used, the measures of the variables, seem valid and reliable? How did the authors attempt to establish this? Could any more have been done in the study to establish measurement validity? (Chapter 4)
11. Was a sample or the entire population of elements used in the study? What type of sample was selected? Was a probability sampling method used? Did the authors think the sample they obtained was generally representative of the population from which it was drawn? Do you? How would you evaluate the likely generalizability of the findings to other populations? (Chapter 5)
12. Was the response rate or participation rate reported? Does it appear likely that those who did not respond or participate were markedly different from those who did participate? Why or why not? Did the author(s) adequately discuss this issue? (Chapter 5)
13. What were the units of analysis? Were they appropriate for the research question? If groups were the units of analysis, were any statements made at any point that are open to the ecological fallacy? If individuals were the units of analysis, were any statements made at any point that suggest reductionist reasoning? (Chapter 6)

14. Was the study design cross-sectional or longitudinal, or did it use both types of data? If the design was longitudinal, what type of longitudinal design was it? Could the longitudinal design have been improved in any way, as by collecting panel data rather than trend data or by decreasing the dropout rate in a panel design? If cross-sectional data were used, could the research question have been addressed more effectively with longitudinal data? (Chapter 6)

15. Were any causal assertions made or implied in the hypotheses or in subsequent discussions? What approach was used to demonstrate the existence of causal effects? Were all three criteria and two cautions for establishing causal relationships addressed? What, if any, variables were controlled in the analysis to reduce the risk of spurious relationships? Should any other variables have been measured and controlled? How satisfied are you with the internal validity of the conclusions? What about external validity? (Chapter 6)

16. Which type of research design was used: experimental, survey, participant observation, historical comparative, or some other? How well was this design suited to the research question posed and the specific hypotheses tested, if any? Why do you suppose the author(s) chose this particular design? How was the design modified in response to research constraints? How was it modified to take advantage of research opportunities? (Chapters 7, 8, 10, 13, 14, 15)

17. Were mixed methods used? What methods were combined, and how can they be distinguished in priority and sequence? In what ways did the combination of qualitative and quantitative data enrich the study's value? (Chapter 12)

18. Was this an evaluation research project? If so, which type of evaluation was it? Which design alternatives did it use? (Chapter 13)

19. Was a secondary data analysis design used? If so, what were the advantages and disadvantages of using data collected in another project? Were Big Data analyzed? If so, did the methods used generate any ethical concerns? (Chapter 14)

20. Was a historical comparative design used? Which type was it? Did the authors address problems resulting from the use of historical or cross-national data? Was a content analysis used? If so, was adequate information provided about each step in the analysis? (Chapter 15)

21. Was any attention given to social context and subjective meanings? If so, what did this add? If not, would it have improved the study? Explain. (Chapter 10)

22. Summarize the findings. How clearly were statistical and/or qualitative data presented and discussed? Were the results substantively important? If meta-analysis was used, was the selection of studies comprehensive and were clear generalizations possible from the population of studies? (Chapters 9, 11)

23. Did the author(s) adequately represent the findings in the discussion or conclusions sections? Were conclusions well grounded in the findings? Are any other interpretations possible? (Chapter 16)

24. Compare the study to others addressing the same research question. Did the study yield additional insights? In what ways was the study design more or less adequate than the design of previous research? (Chapters 2, 16)

25. What additional research questions and hypotheses are suggested by the study's results? What light did the study shed on the theoretical framework used? On social policy questions? (Chapters 2, 13, 16)

*See Chapter 2,Reviewing Prior Research section, under the heading "Single Article Reviews: Aggressive Policing and Educational Performance of Minority Youth," for an example of an article review based on these questions.

APPENDIX B: TABLE OF RANDOM NUMBERS

Line/ Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1	10480	15011	01536	02011	81647	91646	69179	14194	62590	36207	20969	99570	91291	90700
2	22368	46573	25595	85393	30995	89198	27982	53402	93965	34095	52666	19174	39615	99505
3	24130	48360	22527	97265	76393	64809	15179	24830	49340	32081	30680	19655	63348	58629
4	42167	93093	06243	61680	07856	16376	39440	53537	71341	57004	00849	74917	97758	16379
5	37570	39975	81837	16656	06121	91782	60468	81305	49684	60672	14110	06927	01263	54613
6	77921	06907	11008	42751	27756	53498	18602	70659	90655	15053	21916	81825	44394	42880
7	99562	72905	56420	69994	98872	31016	71194	18738	44013	48840	63213	21069	10634	12952
8	96301	91977	05463	07972	18876	20922	94595	56869	69014	60045	18425	84903	42508	32307
9	89579	14342	63661	10281	17453	18103	57740	84378	25331	12566	58678	44947	05585	56941
10	85475	36857	43342	53988	53060	59533	38867	62300	08158	17983	16439	11458	18593	64952
11	28918	69578	88231	33276	70997	79936	56865	05859	90106	31595	01547	85590	91610	78188
12	63553	40961	48235	03427	49626	69445	18663	72695	52180	20847	12234	90511	33703	90322
13	09429	93969	52636	92737	88974	33488	36320	17617	30015	08272	84115	27156	30613	74952
14	10365	61129	87529	85689	48237	52267	67689	93394	01511	26358	85104	20285	29975	89868
15	07119	97336	71048	08178	77233	13916	47564	81056	97735	85977	29372	74461	28551	90707
16	51085	12765	51821	51259	77452	16308	60756	92144	49442	53900	70960	63990	75601	40719
17	02368	21382	52404	60268	89368	19885	55322	44819	01188	65255	64835	44919	05944	55157
18	01011	54092	33362	94904	31273	04146	18594	29852	71585	85030	51132	01915	92747	64951
19	52162	53916	46369	58586	23216	14513	83149	98736	23495	64350	94738	17752	35156	35749
20	07056	97628	33787	09998	42698	06691	76988	13602	51851	46104	88916	19509	25625	58104

Line/ Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
21	48663	91245	85828	14346	09172	30168	90229	04734	59193	22178	30421	61666	99904	32812
22	54164	58492	22421	74103	47070	25306	76468	26384	58151	06646	21524	15227	96909	44592
23	32639	32363	05597	24200	13363	38005	94342	28728	35806	06912	17012	64161	18296	22851
24	29334	27001	87637	87308	58731	00256	45834	15398	46557	41135	10367	07684	36188	18510
25	02488	33062	28834	07351	19731	92420	60952	61280	50001	67658	32586	86679	50720	94953
26	81525	72295	04839	96423	24878	82651	66566	14778	76797	14780	13300	87074	79666	95725
27	29676	20591	68086	26432	46901	20849	89768	81536	86645	12659	92259	57102	80428	25280
28	00742	57392	39064	66432	84673	40027	32832	61362	98947	96067	64760	64584	96096	98253
29	05366	04213	25669	26422	44407	44048	37937	63904	45766	66134	75470	66520	34693	90449
30	91921	26418	64117	94305	26766	25940	39972	22209	71500	64568	91402	42416	07844	69618
31	00582	04711	87917	77341	42206	35126	74087	99547	81817	42607	43808	76655	62028	76630
32	00725	69884	62797	56170	86324	88072	76222	36086	84637	93161	76038	65855	77919	88006
33	69011	65797	95876	55293	18988	27354	26575	08625	40801	59920	29841	80150	12777	48501
34	25976	57948	29888	88604	67917	48708	18912	82271	65424	69774	33611	54262	85963	03547
35	09763	83473	73577	12908	30883	18317	28290	35797	05998	41688	34952	37888	38917	88050
36	91567	42595	27958	30134	04024	86385	29880	99730	55536	84855	29080	09250	79656	73211
37	17955	56349	90999	49127	20044	59931	06115	20542	18059	02008	73708	83317	36103	42791
38	46503	18584	18845	49618	02304	51038	20655	58727	28168	15475	56942	53389	20562	87338
39	92157	89634	94824	78171	84610	82834	09922	25417	44137	48413	25555	21246	35509	20468
40	14577	62765	35605	81263	39667	47358	56873	56307	61607	49518	89656	20103	77490	18062
41	98427	07523	33362	64270	01638	92477	66969	98420	04880	45585	46565	04102	46880	45709
42	34914	63976	88720	82765	34476	17032	87589	40836	32427	70002	70663	88863	77775	69348
43	70060	28277	39475	46473	23219	53416	94970	25832	69975	94884	19661	72828	00102	66794

(*Continued*)

(*Continued*)

Line/ Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
44	53976	54914	06990	67245	68350	82948	11398	42878	80287	88267	47363	46634	06541	97809
45	76072	29515	40980	07391	58745	25774	22987	80059	39911	96189	41151	14222	60697	59583
46	90725	52210	83974	29992	65831	38857	50490	83765	55657	14361	31720	57375	56228	41546
47	64364	67412	33339	31926	14883	24413	59744	92351	97473	89286	35931	04110	23726	51900
48	08962	00358	31662	25388	61642	34072	81249	35648	56891	69352	48373	45578	78547	81788
49	95012	68379	93526	70765	10593	04542	76463	54328	02349	17247	28865	14777	62730	92277
50	15664	10493	20492	38391	91132	21999	59516	81652	27195	48223	46751	22923	32261	85653
51	16408	81899	04153	53381	79401	21438	83035	92350	36693	31238	59649	91754	72772	02338
52	18629	81953	05520	91962	04739	13092	97662	24822	94730	06496	35090	04822	86772	98289
53	73115	35101	47498	87637	99016	71060	88824	71013	18735	20286	23153	72924	35165	43040
54	57491	16703	23167	49323	45021	33132	12544	41035	80780	45393	44812	12515	98931	91202
55	30405	83946	23792	14422	15059	45799	22716	19792	09983	74353	68668	30429	70735	25499
56	16631	35006	85900	98275	32388	52390	16815	69298	82732	38480	73817	32523	41961	44437
57	96773	20206	42559	78985	05300	22164	24369	54224	35083	19687	11052	91491	60383	19746
58	38935	64202	14349	82674	66523	44133	00697	35552	35970	19124	63318	29686	03387	59846
59	31624	76384	17403	53363	44167	64486	64758	75366	76554	31601	12614	33072	60332	92325
60	78919	19474	23632	27889	47914	02584	37680	20801	72152	39339	34806	08930	85001	87820
61	03931	33309	57047	74211	63445	17361	62825	39908	05607	91284	68833	25570	38818	46920
62	74426	33278	43972	10119	89917	15665	52872	73823	73144	88662	88970	74492	51805	99378
63	09066	00903	20795	95452	92648	45454	09552	88815	16553	51125	79375	97596	16296	66092
64	42238	12426	87025	14267	20979	04508	64535	31355	86064	29472	47689	05974	52468	16834
65	16153	08002	26504	41744	81959	65642	74240	56302	00033	67107	77510	70625	28725	34191
66	21457	40742	29820	96783	29400	21840	15035	34537	33310	06116	95240	15957	16572	06004

Line/ Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
67	21581	57802	02050	89728	17937	37621	47075	42080	97403	48626	68995	43805	33386	21597
68	55612	78095	83197	33732	05810	24813	86902	60397	16489	03264	88525	42786	05269	92532
69	44657	66999	99324	51281	84463	60563	79312	93454	68876	25471	93911	25650	12682	73572
70	91340	84979	46949	81973	37949	61023	43997	15263	80644	43942	89203	71795	99533	50501
71	91227	21199	31935	27022	84067	05462	35216	14486	29891	68607	41867	14951	91696	85065
72	50001	38140	66321	19924	72163	09538	12151	06878	91903	18749	34405	56087	82790	70925
73	65390	05224	72958	28609	81406	39147	25549	48542	42627	45233	57202	94617	23772	07896
74	27504	96131	83944	41575	10573	08619	64482	73923	36152	05184	94142	25299	84387	34925
75	37169	94851	39117	89632	00959	16487	65536	49071	39782	17095	02330	74301	00275	48280
76	11508	70225	51111	38351	19444	66499	71945	05422	13442	78675	84081	66938	93654	59894
77	37449	30362	06694	54690	04052	53115	62757	95348	78662	11163	81651	50245	34971	52924
78	46515	70331	85922	38329	57015	15765	97161	17869	45349	61796	66345	81073	49106	79860
79	30986	81223	42416	58353	21532	30502	32305	86482	05174	07901	54339	58861	74818	46942
80	63798	64995	46583	09765	44160	78128	83991	42865	92520	83531	80377	35909	81250	54238
81	82486	84846	99254	67632	43218	50076	21361	64816	51202	88124	41870	52689	51275	83556
82	21885	32906	92431	09060	64297	51674	64126	62570	26123	05155	59194	52799	28225	85762
83	60336	98782	07408	53458	13564	59089	26445	29789	85205	41001	12535	12133	14645	23541
84	43937	46891	24010	25560	86355	33941	25786	54990	71899	15475	95434	98227	21824	19585
85	97656	63175	89303	16275	07100	92063	21942	18611	47348	20203	18534	03862	78095	50136
86	03299	01221	05418	38982	55758	92237	26759	86367	21216	98442	08303	56613	91511	75928
87	79626	06486	03574	17668	07785	76020	79924	25651	83325	88428	85076	72811	22717	50585
88	85636	68335	47539	03129	65651	11977	02510	26113	99447	68645	34327	15152	55230	93448
89	18039	14367	61337	06177	12143	46609	32989	74014	64708	00533	35398	58408	13261	47908

(*Continued*)

(*Continued*)

Line/ Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
90	08362	15656	60627	36478	65648	16764	53412	09013	07832	41574	17639	82163	60859	75567
91	79556	29068	04142	16268	15387	12856	66227	38358	22478	73373	88732	09443	82558	05250
92	92608	82674	27072	32534	17075	27698	98204	63863	11951	34648	88022	56148	34925	57031
93	23982	25835	40055	67006	12293	02753	14827	22235	35071	99704	37543	11601	35503	85171
94	09915	96306	05908	97901	28395	14186	00821	80703	70426	75647	76310	88717	37890	40129
95	50937	33300	26695	62247	69927	76123	50842	43834	86654	70959	79725	93872	28117	19233
96	42488	78077	69882	61657	34136	79180	97526	43092	04098	73571	80799	76536	71255	64239
97	46764	86273	63003	93017	31204	36692	40202	35275	57306	55543	53203	18098	47625	88684
98	03237	45430	55417	63282	90816	17349	88298	90183	36600	78406	06216	95787	42579	90730
99	86591	81482	52667	61583	14972	90053	89534	76036	49199	43716	97548	04379	46370	28672
100	38534	01715	94964	87288	65680	43772	39560	12918	86537	62738	19636	51132	25739	56947

GLOSSARY

Abductive analysis. A qualitative data analysis approach in which researchers produce theoretical hunches for unexpected research findings and develop them with a systematic analysis of variation across a study in relation to a broad range of theories and interchange with other scholars.

Adaptive research design. A research design that develops as the research progresses.

Alternate-forms reliability. A procedure for testing the reliability of responses to survey questions in which subjects' answers are compared after the subjects have been asked slightly different versions of the questions or when randomly selected halves of the sample have been administered slightly different versions of the questions.

Anomalous findings. Unexpected patterns in data.

Anonymity. A provision of research, in which no identifying information is recorded that could be used to link respondents to their responses.

Applied research. Research conducted using the scientific method and addressing immediate, practical concerns, such as determining whether one program or policy has a more desirable impact than another.

Assignment variable. The variable used to specify a cutoff score for eligibility for some treatment in a regression–discontinuity design.

Association. A criterion for establishing a nomothetic causal relationship between two variables. Variation in one variable is related to variation in another variable.

Authenticity. When the understanding of a social process or social setting is one that reflects fairly the various perspectives of participants in that setting.

Availability sampling. A nonprobability sampling method in which elements are selected on the basis of convenience.

Back matter. The section of an applied research report that may include appendixes, tables, figures, and the research instrument(s).

Balanced response choices. When a fixed-choice survey question has an equal number of responses to express positive and negative choices in comparable language.

Bar chart. A graphic for qualitative variables in which the variable's distribution is displayed with solid bars separated by spaces.

Base number (*N*). The total number of cases in a distribution.

Basic science. Research conducted using the scientific method and having the goal of figuring out what the world is like and why it works as it does.

Before-and-after design. A quasi-experimental design consisting of several before-after comparisons involving the same variables but no comparison group.

Behavior coding. An observation in which the researcher categorizes, according to strict rules, the number of times certain behaviors occur.

Belmont Report. Guidelines developed by the U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in 1979 for the protection of human subjects.

Beneficence. The ethical requirement of minimizing possible harms and maximizing benefits in research involving human subjects that was included in the *Belmont Report.*

Big Data. Massive data sets accessible in computer-readable form that are produced by people, available to social scientists, and manageable with today's computers.

Bimodal. A distribution that has two nonadjacent categories with about the same number of cases, and these categories have more cases than any others.

Bipolar response options. Response choices to a survey question that include a middle category and parallel responses with positive and negative valence (can be labeled or unlabeled).

Black box evaluation. The type of evaluation that occurs when an evaluation of program outcomes ignores, and does not identify, the process by which the program produced the effect.

Case study. A setting or group that the analyst treats as an integrated social unit that must be studied holistically and in its particularity.

Case-oriented research. Research that focuses attention on the nation or other unit as a whole.

Case-oriented understanding. An understanding of social processes in a group, formal organization, community, or other collectivity that reflects accurately the standpoint of participants.

Causal effect (idiographic perspective). When a series of concrete events, thoughts, or actions results in a particular event or individual outcome.

Causal effect (nomothetic perspective). When variation in one phenomenon, an independent variable, leads to or results, on average, in variation in another phenomenon, the dependent variable.

Causal validity (internal validity). Exists when a conclusion that A leads to or results in B is correct.

Census. Research in which information is obtained through responses from or information about all available members of an entire population.

Central tendency. The most common value (for variables measured at the nominal level) or the value around which cases tend to center (for a quantitative variable).

Certificate of Confidentiality. A certificate issued to a researcher by the National Institutes of Health that ensures the right to protect from legal subpoenas information obtained about high-risk populations or behaviors—except child abuse or neglect.

Ceteris paribus. Latin phrase meaning "other things being equal."

Chi-square. An inferential statistic used to test hypotheses about relationships between two or more variables in a cross-tabulation.

Closed-ended (fixed-choice) question. A survey question that provides preformatted response choices for the respondent to circle or check.

Cluster. A naturally occurring, mixed aggregate of elements of the population.

Code of Ethics. The professional code of the American Sociological Association for the treatment of human subjects by members, employees, and students; designed to comply with federal policy and revised in 1997.

Coding. The process of assigning a unique numerical code to each response to survey questions.

Cognitive interview. A technique for evaluating questions in which researchers ask people test questions and then probe with follow-up questions to learn how they understood the question and what their answers mean.

Cohort. Individuals or groups with a common starting point. Examples include college class of 1997, people who graduated from high school in the 1980s, General Motors employees who started work between the years 1990 and 2000, and people who were born in the late 1940s or the 1950s (the baby boom generation) or between 1981 and 1996 (millenials).

Combined frequency display. A table that presents together the distributions for a set of conceptually similar variables having the same response categories; common headings are used for the responses.

Comparative historical research. Research comparing data from more than one period in more than one nation.

Comparison group. In an experiment, a group that has been exposed to a different treatment (or value of the independent variable) than the experimental group.

Compensatory rivalry (John Henry effect). A type of contamination in experimental and quasi-experimental designs that occurs when control group members are aware that they are being denied some advantages and increase their efforts by way of compensation.

Complete (or covert) participant. A role in field research in which the researcher does not reveal his or her identity as a researcher to those who are observed while participating.

Complete (or overt) observer. A role in participant observation in which the researcher does not participate in group activities and is publicly defined as a researcher.

Complex mixed-methods design. Qualitative and quantitative methods are combined in a research project that uses more than one of the four basic types of mixed-methods designs or that repeats at least one of those basic types.

Compressed frequency display. A table that presents cross-classification data efficiently by eliminating unnecessary percentages, such as the percentage corresponding to the second value of a dichotomous variable.

Computer-assisted personal interview (CAPI). A personal interview in which a laptop computer is used to display interview questions and to process responses that the interviewer types in, as well as to check that these responses fall within allowed ranges.

Computer-assisted qualitative data analysis. The use of special computer software to assist qualitative analyses through creating, applying, and refining categories; tracing linkages between concepts; and making comparisons between cases and events.

Computer-assisted telephone interview (CATI). A telephone interview in which a questionnaire is programmed into a computer, along with relevant skip patterns, and only valid entries are allowed; incorporates the tasks of interviewing, data entry, and some data cleaning.

Concept. A mental image that summarizes a set of similar observations, feelings, or ideas.

Conceptualization. The process of specifying what we mean by a term. In deductive research, conceptualization helps translate portions of an abstract theory into specific variables that can be used in testable hypotheses. In inductive research, conceptualization is an important part of the process used to make sense of related observations.

Concurrent validity. The type of validity that exists when scores on a measure are closely related to scores on a criterion measured at the same time.

Confidentiality. A provision of research, in which identifying information that could be used to link respondents to their responses is available only to designated research personnel for specific research needs.

Conflict of interest. When a researcher has a significant financial stake in the design or outcome of his or her own research.

Conflict theory. A social theory that identifies conflict between social groups as the primary force in society; understanding the bases and consequences of the conflict is key to understanding social processes.

Conjunctural research. Research that considers the complex combinations in which causal influences operate.

Constant. A characteristic or value that does not change in a situation being studied.

Construct validity. The type of validity that is established by showing that a measure is related to other measures as specified in a theory.

Constructivism. A methodology based on questioning belief in an external reality; emphasizes the importance of exploring the way in which different stakeholders in a social setting construct their beliefs.

Contamination. A source of causal invalidity that occurs when either the experimental or the comparison group is aware of the other group and is influenced in the posttest as a result.

Content analysis. A research method for systematically analyzing and making inferences from recorded human communication, including books, articles, poems, constitutions, speeches, and songs.

Content validity. The type of validity that exists when the full range of a concept's meaning is covered by the measure.

Context effects. When one or more survey questions influence how subsequent questions are interpreted.

Context. A set of interrelated circumstances that alters a relationship between other variables or social processes.

Contextual effect. A variation in relationships of dependent with independent variables between geographic units or other social settings.

Contingent question. A question that is asked of only a subset of survey respondents.

Continuous measure. A measure with numbers indicating the values of variables as points on a continuum.

Control group. A comparison group that receives no treatment.

Convergent validity. An approach to construct validation; the type of validity achieved when one measure of a concept is associated with different types of measures of the same concept.

Conversation analysis. A qualitative method for analyzing the sequential organization and details of ordinary conversation.

Correlation coefficient. A summary statistic that varies from 0 to 1 or –1, with 0 indicating the absence of a linear relationship between two quantitative variables and 1 or –1 indicating that the relationship is completely described by the line representing the regression of the dependent variable on the independent variable.

Correlational analysis. A statistical technique that summarizes the strength of a relationship between two quantitative variables in terms of its adherence to a linear pattern.

Cost–benefit analysis. A type of evaluation research that compares program costs with the economic value of program benefits.

Cost–effectiveness analysis. A type of evaluation research that compares program costs with actual program outcomes.

Counterfactual. The situation that would have occurred if the subjects who were exposed to the treatment were not exposed but otherwise had had identical experiences to those they underwent during the experiment.

Cover letter. The letter sent with a mailed questionnaire that explains the survey's purpose and auspices and encourages the respondent to participate.

Covert observer. A role in participant observation in which the researcher does not participate in group activities and is not publicly defined as a researcher.

Criterion validity. The type of validity that is established by comparing the scores obtained on the measure being validated with those obtained with a more direct or already validated measure of the same phenomenon (the criterion).

Cronbach's alpha. A statistic commonly used to measure interitem reliability.

Cross-population generalizability (external validity). Exists when findings about one group, population, or setting hold true for other groups, populations, or settings.

Cross-sectional comparative research. Research comparing data from one period between two or more nations.

Cross-sectional research design. A study in which data are collected at only one point in time.

Cross-tabulation (crosstab) or contingency table. In the simplest case, a bivariate (two variable) distribution, showing the distribution of one variable for each category of another variable; can be elaborated using three or more variables.

Curvilinear. Any pattern of association between two quantitative variables that does not involve a regular increase or decrease.

Data cleaning. The process of checking data for errors after the data have been entered in a computer file.

Data entry. The process of typing (word processing) or otherwise transferring data on questionnaires or other instruments into a computer file.

Data visualization. A rapidly developing field that focuses on how to represent and present data to facilitate understanding.

Debriefing. A researcher's informing subjects after an experiment about the experiment's purposes and methods and evaluating subjects' personal reactions to the experiment.

Deception. Used in social experiments to create more "realistic" treatments in which the true purpose of the research is not disclosed to participants, often within the confines of a laboratory.

Deductive research. The type of research in which a specific expectation is deduced from a general premise and is then tested.

Demography. The statistical and mathematical study of the size, composition, and spatial distribution of human populations and how these features change over time.

Demoralization. A type of contamination in experimental and quasi-experimental designs that occurs when control group members feel they have been left out of some valuable treatment and perform worse as a result.

Dependent variable. A variable that is hypothesized to vary depending on, or under the influence of, another variable.

Descriptive research. Research in which social phenomena are defined and described.

Descriptive statistics. Statistics used to describe the distribution of and relationship between variables.

Dichotomy. Variable having only two values.

Differential attrition (mortality). A problem that occurs in experiments when comparison groups become different because subjects are more likely to drop out of one of the groups for various reasons.

Digital ethnography. The use of ethnographic methods to study online communities; also termed *netnography, cyberethnography* and *virtual ethnography.*

Digital methods are research strategies for dealing with the ephemeral and unstable nature of online data.

Direction of association. A pattern in a relationship between two variables—the values of variables tend to change consistently in relation to change on the other variable; the direction of association can be either positive or negative.

Discrete measure. A measure that classifies cases in distinct categories.

Discriminant validity. An approach to construct validation; the scores on the measure to be validated are compared with scores on another measure of the same variable and to scores on variables that measure different but related concepts. Discriminant validity is achieved if the measure to be validated is related most strongly to its comparison measure and less so to the measures of other concepts.

Disproportionate stratified sampling. A sampling method in which elements are selected from strata in different proportions from those that appear in the population.

Double negative. A question or statement that contains two negatives, which can muddy the meaning of the question.

Double-barreled question. A single survey question that actually asks two questions but allows only one answer.

Double-blind procedure. An experimental method in which neither the subjects nor the staff delivering experimental treatments know which subjects are getting the treatment and which are receiving a placebo.

Ecological fallacy. An error in reasoning in which incorrect conclusions about individual-level processes are drawn from group-level data.

Effect size. A standardized measure of association—often the difference between the mean of the experimental group and the mean of the control group on the dependent variable, adjusted for the average variability in the two groups.

Efficiency analysis. A type of evaluation research that compares program costs with program effects. It can be either a cost–benefit analysis or a cost–effectiveness analysis.

Elaboration analysis. The process of introducing a third variable into an analysis to better understand—to elaborate—the bivariate (two-variable) relationship under consideration. Additional control variables also can be introduced.

Electronic survey. A survey that is sent and answered by computer, either through e-mail or on the web.

Elements. The individual members of the population whose characteristics are to be measured.

Embedded mixed-methods design. Qualitative and quantitative methods are used concurrently in the research but one is given priority.

Emergence. The appearance of phenomena at a group level that cannot be explained by the properties of individuals within the group; emergence implies phenomena that are more than "the sum of their parts."

Emic focus. Representing a setting with the participants' terms and from their viewpoint.

Empirical generalization. A statement that describes patterns found in data.

Enumeration units. Units that contain one or more elements and that are listed in a sampling frame.

Ethnography. The study of a culture or cultures that some group of people shares, using participant observation over an extended period.

Etic focus. Representing a setting with the researchers' terms and from their viewpoint.

Evaluability assessment. A type of evaluation research conducted to determine whether it is feasible to evaluate a program's effects within the available time and resources.

Evaluation research. Research that describes or identifies the impact of social policies and programs.

Event-based design (cohort study). A type of longitudinal study in which data are collected at two or more points in time from individuals in a cohort.

Event-structure analysis. A systematic method of developing a causal diagram showing the structure of action underlying some chronology of events; the result is an idiographic causal explanation.

Ex post facto control group design A nonexperimental design in which comparison groups are selected after the treatment, program, or other variation in the independent variable has occurred, but when the participants were able to choose the group in which they participated. Often confused with a quasi-experimental design.

Exhaustive. A question's response choices are exhaustive when they cover all possible responses.

Expectancies of experimental staff. A source of treatment misidentification in experiments and quasi-experiments that occurs when change among experimental subjects results from the positive expectancies of the staff who are delivering the treatment rather than from the treatment itself; also called a *self-fulfilling prophecy.*

Experimental group. In an experiment, the group of subjects that receives the treatment or experimental manipulation.

Explanatory research. Research that seeks to identify causes and effects of social phenomena and to predict how one phenomenon will change or vary in response to variation in some other phenomenon.

Exploratory research. Research that seeks to find out how people get along in the setting under question, what meanings they give to their actions, and what issues concern them.

External events. A source of causal invalidity that occurs when events external to the study influence posttest scores; also called an effect of history effect.

Extraneous variable. A variable that influences both the independent and dependent variables, creating a spurious association between them that disappears when the extraneous variable is controlled.

Face validity. The type of validity that exists when an inspection of items used to measure a concept suggests that they are appropriate "on their face."

Factorial survey. A survey in which randomly selected subsets of respondents are asked different questions, or are asked to respond to different vignettes, to determine the causal effect of the variables represented by these differences.

Federal Policy for the Protection of Human Subjects (also known as the Common Rule). Specific regulations adopted in 1991 by the U.S. Department of Health and Human Services and the U.S. Food and Drug Administration that were based on the principles of the *Belmont Report.* Revised in 2017, implemented in 2019.

Feedback. Information about service delivery system outputs, outcomes, or operations that can guide program input.

Fence-sitters. Survey respondents who see themselves as being neutral on an issue and choose a middle (neutral) response that is offered.

Field experiment. A study using an experimental design that is conducted in a real-world setting.

Field notes. Notes that describe what has been observed, heard, or otherwise experienced in a participant observation study. These notes usually are written after the observational session.

Field research. Research in which natural social processes are studied as they happen and left relatively undisturbed.

Field researcher. A researcher who uses qualitative methods to conduct research in the field.

Filter question. A survey question used to identify a subset of respondents who then are asked other questions.

Fixed-sample panel design (panel study). A type of longitudinal study in which data are collected from the same individuals—the panel—at two or more points in time. In another type of panel design, panel members who leave are replaced with new members.

Floaters. Survey respondents who provide an opinion on a topic in response to a closed-ended question that does not include a "Don't know" option, but who will choose "Don't know" if it is available.

Focus groups. A qualitative method that involves unstructured group interviews in which the focus group leader actively encourages discussion among participants on the topics of interest.

Forced-choice questions. Closed-ended survey questions that do not include "Don't know" as an explicit response choice.

Formative evaluation. Process evaluation that is used to shape and refine program operations.

Frequency distribution. A numerical display showing the number of cases, and usually the percentage of cases (the relative frequencies), corresponding to each value or group of values of a variable

Frequency polygon. A graphic for quantitative variables in which a continuous line connects data points representing the variable's distribution.

Front matter. The section of an applied research report that includes an executive summary, an abstract, and a table of contents.

Functionalism. A social theory that explains social patterns in terms of their consequences for society as a whole and that emphasizes the interdependence of social institutions and their common interest in maintaining the social order.

Gamma. A measure of association that is sometimes used in crosstabular analysis.

Gatekeeper. A person in a field setting who can grant researchers access to the setting.

Generalizability. Exists when a conclusion holds true for the population, group, setting, or event that we say it does, given the conditions that we specify.

Grand tour question. A broad question at the start of an interview that seeks to engage the respondent in the topic of interest.

Grounded theory. A systematic theory developed inductively, based on observations that are summarized into conceptual categories, reevaluated in the research setting, and gradually refined and linked to other conceptual categories.

Group-administered survey. A survey that is completed by individual respondents who are assembled in a group.

Hawthorne effect. A type of contamination in research designs that occurs when members of the treatment group change relative to the dependent variable because their participation in the study makes them feel special.

Health Insurance Portability and Accountability Act (HIPAA). Congressional legislation passed in 1996 that creates stringent regulations for the protection of health care data.

Hermeneutic circle. A representation of the dialectical process in which the researcher obtains information from multiple stakeholders in a setting, refines his or her understanding of the setting, and then tests that understanding with successive respondents.

Histogram. A graphic for quantitative variables in which the variable's distribution is displayed with adjacent bars.

Historical events research. Research in which social events are studied at one past time.

Historical process research. Research in which historical processes are studied over a long time.

History effect. A source of causal invalidity that occurs when events external to the study influence posttest scores; also called an effect of external events.

Holistic research. Research concerned with the context in which events occurred and the interrelations between different events and processes.

Hypothesis. A tentative statement about empirical reality, involving a relationship between two or more variables.

Idiographic causal explanation. An explanation that identifies the concrete, individual sequence of events, thoughts, or actions that resulted in a particular outcome for a particular individual or that led to a particular event; may be termed an *individualist* or *historicist* explanation.

Idiosyncratic (or "random") errors. Errors that affect individuals or other cases in unique ways that are unlikely to be repeated in just the same way.

Idiosyncratic variation. Variation in responses to questions that is caused by individuals' reactions to particular words or ideas in the question instead of by variation in the concept that the question is intended to measure.

Illogical reasoning. When we prematurely jump to conclusions or argue on the basis of invalid assumptions.

Impact evaluation (or analysis). Analysis of the extent to which a treatment or other service has an effect; also known as **summative evaluation**.

Inaccurate observation. An observation based on faulty perceptions of empirical reality.

Independent variable. A variable that is hypothesized to cause, or lead to, variation in another variable.

Index. The sum or average of responses to a set of questions about a concept.

Indicator. The question or other operation used to indicate the value of cases on a variable.

Inductive research. The type of research in which general conclusions are drawn from specific data.

Inferential statistics. A mathematical tool for estimating how likely it is that a statistical result based on data from a random sample is representative of the population from which the sample is assumed to have been selected.

In-person interview. A survey in which an interviewer questions respondents face-toface and records their answers.

Inputs. The resources, raw materials, clients, and staff that go into a program.

Institutional review board (IRB). A group of organizational and community representatives required by federal law to review the ethical issues in all proposed research that is federally funded, involves human subjects, or has any potential for harm to human subjects.

Integrated mixed-methods design. Qualitative and quantitative methods are used concurrently and both are given equal importance.

Integrative approach. An orientation to evaluation research that expects researchers to respond to the concerns of people involved with the program— stakeholders—as well as to the standards and goals of the social scientific community.

Intensive (in-depth) interviewing. A qualitative method that involves open-ended, relatively unstructured questioning in which the interviewer seeks in-depth information on the interviewee's feeling, experiences, and perceptions (Lofland and Lofland 1984.12).

Intent-to-treat analysis. When analysis of the effect of a treatment on outcomes in an experimental design compares outcomes for all those who were assigned to the treatment group with outcomes for all those who were assigned to the control group, whether or not participants remained in the treatment group.

Interactive voice response (IVR). A survey in which respondents receive automated calls and answer questions by pressing numbers on their touch-tone phones or speaking numbers that are interpreted by computerized voice recognition software.

Intercoder reliability. When the same codes are entered by different coders who are recording the same data.

Interitem reliability. An approach that calculates reliability based on the correlation among multiple items used to measure a single concept; also known as internal consistency.

Interobserver reliability. When similar measurements are obtained by different observers rating the same persons, events, or places.

Interpretive questions. Questions included in a questionnaire or interview schedule to help explain answers to other important questions.

Interpretivism. The belief that the subjective meanings people give to their experiences are a key focus for social science research without assuming that social processes can be identified objectively.

Interquartile range. The range in a distribution between the end of the first quartile and the beginning of the third quartile.

Intersubjective agreement. Agreement between scientists about the nature of reality; often upheld as a more reasonable goal for science than certainty about an objective reality.

Inter-university Consortium for Political and Social Research (ICPSR). The academic consortium that archives data sets online from major surveys and other social science research and makes them available for analysis by others.

Interval level of measurement. A measurement of a variable in which the numbers indicating a variable's values represent fixed measurement units but have no absolute, or fixed, zero point.

Interval–ratio level of measurement. A measurement of a variable in which the numbers indicating a variable's values represent fixed measurement units but may not have an absolute, or fixed, zero point.

Interview schedule. The survey instrument containing the questions asked by the interviewer in an in-person or phone survey.

Intrarater (or intraobserver) reliability. Consistency of ratings by an observer of an unchanging phenomenon at two or more points in time.

Jottings. Brief notes written in the field about highlights of an observation period.

Justice. The ethical principle of distributing benefits and risks of research fairly in research involving human subjects that was included in the *Belmont Report.*

Key informant. An insider who is willing and able to provide a field researcher with superior access and information, including answers to questions that arise in the course of the research.

Labeled unipolar response options. Response choices for a survey question that use words to identify categories ranging from low to high (or high to low).

Level of measurement. The mathematical precision with which the values of a variable can be expressed. The nominal level of measurement, which is qualitative, has no mathematical interpretation; the quantitative levels of measurement—ordinal, interval, and ratio—are progressively more precise mathematically.

Likert item. A statement followed by response choices ranging from "strongly agree" to "strongly disagree."

Longitudinal research design. A study in which data are collected that can be ordered in time; also defined as research in which data are collected at two or more points in time.

Mailed survey. A survey involving a mailed questionnaire to be completed by the respondent.

Marginal distribution. The summary distributions in the margins of a cross-tabulation that correspond to the frequency distribution of the row variable and of the column variable.

Matching. A procedure for equating the characteristics of individuals in different comparison groups in an experiment. Matching can be done on either an individual or an aggregate basis. For individual matching, individuals who are similar in key characteristics are paired before assignment, and then the two members of each pair are assigned to the two groups. For aggregate matching, also termed *blocking,* groups that are chosen for comparison are similar in the distribution of key characteristics.

Matrix questions. A series of survey questions that concern a common theme and that have the same response choices.

Matrix. A form on which particular features of multiple cases or instances can be recorded systematically so that a qualitative data analyst can examine them later.

Mean. The arithmetic, or weighted, average, computed by adding the value of all the cases and dividing by the total number of cases.

Measure of association. A type of descriptive statistic that summarizes the strength of an association.

Measurement validity. Exists when a measure measures what we think it measures.

Measurement. The process of linking abstract concepts to empirical indicants.

Mechanism. A discernible process that creates a causal connection between two variables.

Median. The position average, or the point that divides a distribution in half (the 50th percentile).

Mediator. A variable involved in a causal mechanism (intervening variable).

Meta-analysis. The quantitative analysis of findings from multiple studies.

Method of agreement. A method proposed by John Stuart Mill for establishing a causal relation, in which the values of cases that agree on an outcome variable also agree on the value of the variable hypothesized to have a causal effect, although they differ on other variables.

Method of difference. A method proposed by John Stuart Mill for establishing a causal relation, in which the values of cases that differ on an outcome variable also differ on the value of the variable hypothesized to have a causal effect, although they agree on other variables.

Milgram's obedience experiments. Experiments begun in 1960 at Yale University by psychologist Stanley Milgram to determine the likelihood of people following orders from an authority despite their own sentiments; widely cited as helping to understand the emergence of phenomena such as Nazism and mass cults.

Mixed methods. Research that combines qualitative and quantitative methods in an investigation of the same or related research question(s).

Mixed-mode survey. A survey that is conducted by more than one method, allowing the strengths of one survey design to compensate for the weaknesses of another and maximizing the likelihood of securing data from different types of respondents; for example, nonrespondents in a mailed survey may be interviewed in person or over the phone.

Mode. The most frequent value in a distribution; also termed the **probability average**.

Moderator. A variable that identifies a context for the effect of other variables.

Monotonic. A pattern of association in which the value of cases on one variable increases or decreases fairly regularly across the categories of another variable.

Multiple group before-and-after design. A type of quasi-experimental design in which several before-and-after comparisons are made involving the same independent and dependent variables but different groups.

Multistage cluster sampling. A sampling method in which elements are selected in two or more stages, with the first stage being the random selection of naturally occurring clusters and the last stage being the random selection of elements within clusters.

Multitrait–multimethod matrix. A method of evaluating the validity of measures by determining whether measuring the same phenomenon with different methods leads to convergent results and measuring different phenomena with the same method leads to divergent results.

Mutually exclusive. A question's response choices are mutually exclusive when every case can be classified as having only one attribute (or value).

Narrative analysis. A form of qualitative analysis in which the analyst focuses on how respondents impose order on the flow of experience in their lives and thus make sense of events and actions in which they have participated.

Narrative explanation. An idiographic causal explanation that involves developing a narrative of events and processes that indicate a chain of causes and effects.

Needs assessment. A type of evaluation research that attempts to determine the needs of some population that might be met with a social program.

Ngrams. Frequency graphs produced by Google's database of all words printed in more than one third of the world's books over time (with coverage still expanding).

Nominal level of measurement. Variables whose values have no mathematical interpretation; they vary in kind or quality, but not in amount.

Nomothetic causal explanation. An explanation that identifies common influences on cases or events.

Nonequivalent control group design. A quasi-experimental design in which experimental and comparison groups are designated before the treatment occurs but are not created by random assignment.

Nonprobability sampling method. A sampling method in which the probability of selection of population elements is unknown.

Nonrespondents. People or other entities who do not participate in a study although they are selected for the sample.

Nonspuriousness. A criterion for establishing a causal relation between two variables; when a relationship between two variables is not caused by variation in a third variable.

Normal distribution. A symmetric, bell-shaped distribution that results from chance variation around a central value.

Normal science. The gradual, incremental research conducted by scientists within the prevailing scientific paradigm.

Nuremberg War Crime Trials. The International Military Tribunal held by the victorious Allies after World War II in Nuremberg, Germany, that exposed the horrific medical experiments conducted by Nazi doctors and others in the name of "science."

Office for Protection From Research Risks, National Institutes of Health. The office in the U.S. Department of Health and Human Services that provides leadership and supervision about the protection of the rights, welfare, and well-being of subjects involved in research conducted or supported by DHHS, including monitoring IRBs.

Omnibus survey. A survey that covers a range of topics of interest to different social scientists.

Open-ended question. A survey question to which the respondent replies in his or her own words, either by writing or by talking.

Operationalization. The process of specifying the measures that will indicate the value of cases on a variable.

Oral history. Data collected through intensive interviews with participants in past events.

Ordinal level of measurement. A measurement of a variable in which the numbers indicating a variable's values specify only the order of the cases, permitting *greater than* and *less than* distinctions.

Outcomes. The impact of the program process on the cases processed.

Outlier. An exceptionally high or low value in a distribution.

Outputs. The services delivered or new products produced by the program process.

Overgeneralization. When we unjustifiably conclude that what is true for some cases is true for all cases.

Paradigm wars. The intense debate from the 1970s to the 1990s between social scientists over the value of positivist and interpretivist/constructivist research philosophies; also see *scientific paradigm.*

Participant observation. A qualitative method for gathering data that involves developing a sustained relationship with people while they go about their normal activities.

Participant observer. A researcher who gathers data through participating and observing in a setting where he or she develops a sustained relationship with people while they go about their normal activities. The term *participant observer* is often used to refer to a continuum of possible roles, from complete observation, in which the researcher does not participate along with others in group activities, to complete participation, in which the researcher participates without publicly acknowledging being an observer.

Participatory action research (PAR). A type of research in which the researcher involves members of the population to be studied as active participants throughout the research process, from the selection of a research focus to the reporting of research results and efforts to make changes based on the research; also termed *community-based participatory research.*

Part–whole question effects. When responses to a general or summary survey question about a topic are influenced by responses to an earlier, more specific question about that topic.

Percentages. Relative frequencies, computed by dividing the frequency of cases in a particular category by the total number of cases and then multiplying by 100.

Periodicity. A sequence of elements (in a list to be sampled) that varies in some regular, periodic pattern.

Phone survey. A survey in which interviewers question respondents over the phone and then record their answers.

Photo voice. A method in which research participants take pictures of their everyday surroundings with cameras the researcher distributes, and then meet in a group with the researcher to discuss the pictures' meaning.

Placebo effect. A source of treatment misidentification that can occur when subjects receive a fake "treatment" they think is beneficial and improve because of that expectation even though they did not receive the actual treatment or received a treatment that had no real effect.

Plagiarism. Presenting as one's own the ideas or words of another person or persons for academic evaluation without proper acknowledgment.

Policy research. A process in which research results are used to provide policy actors with recommendations for action that are based on empirical evidence and careful reasoning.

Population parameter. The value of a statistic, such as a mean, computed using the data for the entire population; a sample statistic is an estimate of a population parameter.

Population. The entire set of individuals or other entities to which study findings are to be generalized.

Positivism. The belief, shared by most scientists, that there is a reality that exists quite apart from our own perception of it, that it can be understood through observation, and that it follows general laws.

Postpositivism. A philosophical view that modifies the positivist premise of an external, objective reality by recognizing its complexity, the limitations of human observers, and therefore the impossibility of developing more than a partial understanding of reality.

Posttest. In experimental research, the measurement of an outcome (dependent) variable after an experimental intervention or after a presumed independent variable has changed for some other reason.

Predictive validity. The type of validity that exists when a measure predicts scores on a criterion measured in the future.

Pretest. In experimental research, the measurement of an outcome (dependent) variable before an experimental intervention or change in a presumed independent variable for some other reason. The pretest is the same "test" as the posttest, but it is administered at a different time.

Probability average. The most frequent value in a distribution; also termed the **mode.**

Probability of selection. The likelihood that an element will be selected from the population for inclusion in the sample. In a census of all elements of a population, the probability that a particular element will be selected is 1.0. If half the elements in the population are sampled by chance (say, by tossing a coin), the probability of selection for each element is one half, or .5. As the size of the sample as a proportion of the population decreases, so does the probability of selection.

Probability sampling method. A sampling method that relies on a random, or chance, selection method so that the probability of selection of population elements is known.

Process consent. An interpretation of the ethical standard of voluntary consent that allows participants to change their decision about participating at any point by requiring that the researcher check with participants at each stage of the project about their willingness to continue in the project.

Process evaluation. Evaluation research that investigates the process of service delivery.

Process responsiveness. An ethical standard for research that requires checking with research participants throughout a study to ensure their ongoing comfort with and consent to study participation.

Program process. The complete treatment or service delivered by the program.

Program theory. A descriptive or prescriptive model of how a program operates and produces effects.

Progressive focusing. The process by which a qualitative analyst interacts with the data and gradually refines his or her focus.

Proportionate stratified sampling. A sampling method in which elements are selected from strata in exact proportion to their representation in the population.

Pseudoscience. Claims presented so that they appear scientific even though they lack supporting evidence and plausibility.

Publication bias. The greater likelihood of publication of articles reporting tests of hypotheses that were supported.

Purposive sampling. A nonprobability sampling method in which elements are selected for a purpose, usually because of their unique position.

Qualitative comparative analysis (QCA). A systematic type of qualitative analysis that identifies the combination of factors that had to be present across multiple cases to produce a particular outcome.

Qualitative methods. Methods such as participant observation, intensive interviewing, and focus groups that are designed to capture social life as participants experience it rather than in categories predetermined by the researcher. These methods rely on written or spoken words or observations that do not often have a direct numerical interpretation and typically involve exploratory research questions, inductive reasoning, an orientation to social context and human subjectivity, and the meanings attached by participants to events and to their lives.

Quantitative methods. Methods such as surveys and experiments that record variation in social life in terms of categories that vary in amount. Data that are treated as quantitative are either numbers or attributes that can be ordered by magnitude.

Quartiles. The points in a distribution corresponding to the first 25% of the cases, the first 50% of the cases, and the first 75% of the cases.

Quasi-experimental design. A research design in which there is a comparison group that is comparable to the experimental group in critical ways, but subjects are not randomly assigned to the comparison and experimental groups.

Questionnaire. The survey instrument containing the questions in a self-administered survey.

Quota sampling. A nonprobability sampling method in which elements are selected to ensure that the sample represents certain characteristics in proportion to their prevalence in the population.

Random assignment. A procedure by which each experimental subject or group is placed in a condition randomly.

Random digit dialing. The random machine dialing of numbers within designated phone prefixes, which creates a random sample for phone surveys.

Random number table. A table containing lists of numbers that are ordered solely based on chance; it is used for drawing a random sample.

Random sampling error (chance sampling error). Differences between the population and the sample that are due only to chance factors (random error), not to systematic sampling error. Random sampling error may or may not result in an unrepresentative sample. The magnitude of sampling error resulting from chance factors can be estimated statistically.

Random sampling. A method of sampling that relies on a random, or chance, selection method so that every element of the sampling frame has a known probability of being selected.

Randomization. The random assignment of cases, as by the toss of a coin.

Randomized comparative change design. The classic true experimental design in which subjects are assigned randomly to two groups; both these groups receive a pretest, then one group receives the experimental intervention (the "treatment"), and then both groups receive a posttest. Also known as a *pretest–posttest control group design.*

Randomized comparative posttest design. A true experimental design in which subjects are assigned randomly to two groups—one group then receives the experimental intervention and both groups receive a posttest; there is no pretest. Also known as a *posttest-only control group design.*

Range. The true upper limit in a distribution minus the true lower limit (or the highest rounded value minus the lowest rounded value, plus one).

Ratio level of measurement. A measurement of a variable in which the numbers indicating a variable's values represent fixed measuring units and an absolute zero point.

Rational choice theory. A social theory that explains individual action with the principle that actors choose actions that maximize their gains from taking that action.

Reactive effects. The changes in individual or group behavior that result from being observed or otherwise studied.

Reductionist fallacy (reductionism). An error in reasoning that occurs when incorrect conclusions about group-level processes are based on individual-level data; also known as an individualist fallacy.

Reflexivity. Sensitivity of and adaptation by the researcher to his or her influence in the research setting.

Regression analysis. A statistical technique for characterizing the pattern of a relationship between two quantitative variables in terms of a linear equation and for summarizing the strength of this relationship in terms of its deviation from that linear pattern.

Regression effect. A source of causal invalidity that occurs when subjects who are chosen for a study because of their extreme scores on the dependent variable become less extreme on the posttest because of natural cyclical or episodic change in the variable.

Regression–discontinuity design. A quasi-experimental design in which individuals are assigned to a treatment and a comparison group solely on the basis of a cutoff score on some assignment variable, and then treatment effects are identified by a discontinuity in the regression line that displays the relation between the outcome and the assignment variable at the cutoff score.

Reliability measures. Statistics that summarize the consistency among a set of measures; Cronbach's alpha is the most common measure of the reliability of a set of items included in an index.

Reliability. When a measurement procedure yields consistent scores when the phenomenon being measured is not changing.

Repeated cross-sectional design (trend study). A type of longitudinal study in which data are collected at two or more points in time from different samples of the same population.

Repeated measures panel design. A quasi-experimental design consisting of several pretest and posttest observations of the same group.

Replacement sampling. A method of sampling in which sample elements are returned to the sampling frame after being selected, so they may be sampled again. Random samples may be selected with or without replacement.

Replications. Repetitions of a study using the same research methods to answer the same research question.

Representative sample. A sample that "looks like" the population from which it was selected in all respects that are potentially relevant to the study. The distribution of characteristics among the elements of a representative sample is the same as the distribution of those characteristics among the total population. In an unrepresentative sample, some characteristics are overrepresented or underrepresented.

Research circle. A diagram of the elements of the research process, including theories, hypotheses, data collection, and data analysis.

Research program with mixed methods. Qualitative and quantitative methods are used in multiple studies in sequence and are given equal priority.

Resistance to change. The reluctance to change our ideas in light of new information.

Respect for persons. The ethical principle of treating persons as autonomous agents and protecting those with diminished autonomy in research involving human subjects that was included in the *Belmont Report.*

Reverse outlining. Outlining the sections in an already written draft of a paper or report to improve its organization in the next draft.

Sample generalizability. Exists when a conclusion based on a sample, or subset, of a larger population holds true for that population.

Sample statistic. The value of a statistic, such as a mean, computed from sample data.

Sample. A subset of a population that is used to study the population as a whole.

Sampling error. Any difference between the characteristics of a sample and the characteristics of a population; the larger the sampling error, the less representative the sample.

Sampling frame. A list of all elements or other units containing the elements in a population.

Sampling interval. The number of cases from one sampled case to another in a systematic random sample.

Sampling units. Units listed at each stage of a multistage sampling design.

Saturation point. The point at which subject selection is ended in intensive interviewing, when new interviews seem to yield little additional information.

Science. A set of logical, systematic, documented methods for investigating nature and natural processes; the knowledge produced by these investigations.

Scientific paradigm. A set of beliefs that guide scientific work in an area, including unquestioned presuppositions, accepted theories, and exemplary research findings.

Secondary data analysis. The method of using preexisting data in a different way or to answer a different research question than intended by those who collected the data.

Secondary data. Previously collected data that are used in a new analysis.

Selection bias. A source of internal (causal) invalidity that occurs when characteristics of experimental and comparison group subjects differ in any way that influences the outcome.

Selective distribution of benefits. An ethical issue about how much researchers can influence the benefits subjects receive as part of the treatment being studied in a field experiment.

Selective observation. Choosing to look only at things that are in line with our preferences or beliefs.

Serendipitous findings. Unexpected patterns in data, which stimulate new explanations, insights, or theoretical approaches.

Simple random sampling. A method of sampling in which every sample element is selected only based on chance, through a random process.

Skewness. The extent to which cases are clustered more at one or the other end of the distribution of a quantitative variable rather than in a symmetric pattern around its center. Skew can be positive (a right skew), with the number of cases tapering off in the positive direction, or negative (a left skew), with the number of cases tapering off in the negative direction.

Skip pattern. The unique combination of questions created in a survey by filter questions and contingent questions.

Snowball sampling. A nonprobability sampling method in which sample elements are selected as they are identified by successive informants or interviewees.

Social desirability bias. The tendency to "agree" with a statement just to avoid seeming disagreeable.

Social research question. A question about the social world that is answered through the collection and analysis of firsthand, verifiable, empirical data.

Social science (researcher) approach. An orientation to evaluation research that expects researchers to emphasize the importance of researcher expertise and maintenance of autonomy from program stakeholders.

Social science. The use of scientific methods to investigate individuals, societies, and social processes; the knowledge produced by these investigations.

Solomon four-group design. A type of experimental design that combines a randomized pretest–posttest control group design with a randomized posttest-only design, resulting in two experimental groups and two comparison groups.

Specification. A type of relationship involving three or more variables in which the association between the independent and dependent variables varies across the categories of one or more other control variables.

Split-ballot design. Unique questions or other modifications in a survey administered to randomly selected subsets of the total survey sample, so that more questions can be included in the entire survey or so that responses to different question versions can be compared.

Split-half reliability. Reliability achieved when responses to the same questions by two randomly selected halves of a sample are about the same.

Spurious relationship. A relationship between two variables that is caused by variation in a third variable.

Staged mixed-methods design. Qualitative and quantitative methods are used in sequence in the research and one is given priority.

Stakeholder approach. An orientation to evaluation research that expects researchers to be responsive primarily to the people involved with the program; also termed **responsive evaluation.**

Stakeholders. Individuals and groups who have some basis of concern with the program.

Standard deviation. The square root of the average squared deviation of each case from the mean.

Statistical control. A method in which one variable is held constant so that the relationship between two (or more) other variables can be assessed without the influence of variation in the control variable.

Statistical significance. The mathematical likelihood that an association is due to chance, judged by a criterion set by the analyst (often that the probability is less than 5 out of 100 or $p < .05$).

Stratified random sampling. A sampling method in which sample elements are selected separately from population strata that are identified in advance by the researcher.

Street-level bureaucrats. Officials who serve clients and have a high degree of discretion.

Subject fatigue. Problems caused by panel members growing weary of repeated interviews and dropping out of a study or becoming so used to answering the standard questions in the survey that they start giving stock or thoughtless answers.

Subtables. Tables describing the relationship between two variables within the discrete categories of one or more other control variables.

Summative evaluation: Analysis of the extent to which a treatment or other service has an effect; also known as impact analysis or impact evaluation.

Survey pretest. A method of evaluating survey questions and procedures by testing them on a small sample of individuals like those to be included in the actual survey and then reviewing responses to the questions and reactions to the survey procedures.

Survey research. Research in which information is obtained from a sample of individuals through their responses to questions about themselves or others.

Symbolic interaction theory. A social theory that focuses on the symbolic nature of social interaction—how social interaction conveys meaning and promotes socialization.

Systematic bias. Overrepresentation or underrepresentation of some population characteristics in a sample resulting from the method used to select the sample; a sample shaped by systematic sampling error is a biased sample.

Systematic errors. Errors that result from factors that are not what an instrument is intended to measure and that affect individuals or other cases in ways that are likely to recur in just the same way, thus creating measurement bias.

Systematic observation. A strategy that increases the reliability of observational data by using explicit rules that standardize coding practices across observers.

Systematic random sampling. A method of sampling in which sample elements are selected from a list or from sequential files, with every *n*th element being selected after the first element is selected randomly within the first interval.

Systematic review. A literature review that "uses a specific and reproducible method to identify, select and appraise studies of a previously agreed level of quality (either including all studies or only those that pass a minimum quality threshold) that are relevant to a particular question." (Booth et al. 2016.11)

Tacit knowledge. In field research, a credible sense of understanding of social processes that reflects the researcher's awareness of participants' actions as well as their words, and of what they fail to state, feel deeply, and take for granted.

Target population. A set of elements larger than or different from the population sampled and to which the researcher would like to generalize study findings.

Tearoom Trade. The study by sociologist Laud Humphreys of men who engaged in homosexual behavior in public facilities, including subsequent later interviews in their homes after recording their license plate numbers; widely cited in discussions of the need for informed consent to research.

Temporal research. Research that accounts for the related series of events that unfold over time.

Test–retest reliability. A measurement showing that measures of a phenomenon at two points in time are highly correlated, if the phenomenon has not changed or has changed only as much as the measures have changed.

Theoretical sampling. A sampling method recommended for field researchers by Glaser and Strauss (1967). A theoretical sample is drawn in a sequential fashion, with settings or individuals selected for study as earlier observations or interviews indicate that these settings or individuals are influential.

Theory. A logically interrelated set of propositions about empirical reality.

Theory-driven evaluation. A program evaluation that is guided by a theory that specifies the process by which the program has an effect.

Thick description. A rich description that conveys a sense of what it is like from the standpoint of the natural actors in that setting.

Time order. A criterion for establishing a causal relation between two variables; the variation in the presumed cause (the independent variable) must occur before the variation in the presumed effect (the dependent variable).

Time series design. A quasi-experimental design consisting of many pretest and posttest observations of the same group over an extended period.

Triangulation. The use of multiple methods to study one research question; also used to mean the use of two or more different measures of the same variable.

True experiment. An experiment in which subjects are assigned randomly to an experimental group that receives a treatment or other manipulation of the independent variable and a comparison group that does not receive the treatment or receives some other manipulation; outcomes are measured in a posttest.

Tuskegee Study of Untreated Syphilis in the Negro Male. The U.S. Public Health Service study of the "natural" course of syphilis that followed 399 low-income African American men from the 1930s to 1972, without providing them with penicillin after the drug was discovered to treat the illness. The study was stopped after it was exposed in 1972, resulting in an out-of-court settlement and then, in 1997, an official public apology by President Bill Clinton.

Unbalanced response choices. When a fixed-choice survey question has a different number of positive and negative response choices.

Unimodal. A distribution of a variable in which there is only one value that is the most frequent.

Units of analysis. The level of social life on which a research question is focused, such as individuals, groups, towns, or nations.

Units of observation. The cases about which measures actually are obtained in a sample.

Unlabeled unipolar response options. Response choices for a survey question that use numbers to identify categories ranging from low to high (or high to low).

Unobtrusive measure. A measurement based on physical traces or other data that are collected without the knowledge or participation of the individuals or groups that generated the data.

Validity. The state that exists when statements or conclusions about empirical reality are correct.

Variability. The extent to which cases are spread out through the distribution or clustered in just one location.

Variable. A characteristic or property that can vary (take on different values or attributes).

Variable-oriented research. Research that focuses attention on variables representing particular aspects of the cases studied and then examines the relations between these variables across sets of cases.

Variance. A statistic that measures the variability of a distribution as the average squared deviation of each case from the mean.

Video ethnography. The use of audiovisual methods and editing techniques to record, analyze, and present one or more viewable social processes, actions, or events in interpretable segments.

Visual sociology. Sociological research in which the social world is "observed" and interpreted through photographs, films, and other images.

Web survey. A survey that is accessed and responded to on the World Wide Web.

Zimbardo prison simulation study. The famous prison simulation study at Stanford University by psychologist Philip Zimbardo, designed to investigate the impact of social position on behavior— specifically, the impact of being either a guard or a prisoner in a "total institution"; widely cited as demonstrating the likelihood of emergence of sadistic behavior in guards.

BIBLIOGRAPHY

270toWin. n.d. "National 2020 Presidential Election Polls" (https://www.270towin.com/2020-polls-biden-trump/national/).

Abbey, Antonia. 2002. "Alcohol-Related Sexual Assault: A Common Problem Among College Students." *Journal of Studies on Alcohol* (Suppl. 14):118-28.

Abbott, Andrew. 1992. "From Causes to Events: Notes on Narrative Positivism." *Sociological Methods and Research* 20(May):428-55.

Abbott, Andrew. 1994. "History and Sociology: The Lost Synthesis." Pp. 77–112 in *Engaging the Past: The Uses of History Across the Social Sciences,* edited by Eric H. Monkkonen. Durham, NC: Duke University Press.

Abel, David. 2000. "Census May Fall Short at Colleges." *The Boston Sunday Globe,* March 26, B1, B4.

Abernathy, David. 2017. *Using Geodata and Geolocation in the Social Sciences: Mapping Our Connected World.* Thousand Oaks, CA: Sage.

Abma, Tineke A. 2005. "Responsive Evaluation: Its Meaning and Special Contribution to Health Promotion." *Evaluation and Program Planning* 28:279-89.

Abrutyn, Seth, and Anna S. Mueller. 2014. "Are Suicidal Behaviors Contagious in Adolescence? Using Longitudinal Data to Examine Suicide Suggestion." *American Sociological Review* 79:211-27.

Acocella, Ivana and Silvia Cataldi. 2020. *Using Focus Groups.* London: Sage Publications UK.

Adair, John G. Terrance W. Dushenko, and R. C. Lindsay. 1985. "Ethical Regulations and Their Impact on Research Practice." *American Psychologist* 40:59-72.

Aguirre, Adalberto Jr. and David V. Baker. 1993. "Racial Prejudice and the Death Penalty: A Research Note." *Social Justice* 20:150-55.

Aiden, Erez and Jean-Baptiste Michel. 2013. *Uncharted: Big Data as a Lens on Human Culture.* New York, NY: Riverhead Books.

Akyüz, Aygül, Tülay Yavan, Gönül Şahiner, and Ayşe Kılıç. 2012. "Domestic Violence and Woman's Reproductive Health: A Review of the Literature." *Aggression and Violent Behavior* 17:514-18.

Aldrich, James O. and Hilda M. Rodriguez. 2013. *Building SPSS Graphs to Understand Data.* Thousand Oaks, CA: Sage.

Alfred, Randall. 1976. "The Church of Satan." Pp. 180-202 in *The New Religious Consciousness,* edited by Charles Glock and Robert Bellah. Berkeley, CA: University of California Press.

Alise, Mark A. and Charles Teddlie. 2010. "A Continuation of the Paradigm Wars? Prevalence Rates of Methodological Approaches Across the Social/Behavioral Sciences." *Journal of Mixed Methods Research* 4:103-26.

Alkire, Sabina, José Manuel Roche, Maria Emma Santos, and Suman Seth. 2011. "Multidimensional Poverty Index 2011: Brief Methodological Note." Oxford, England: Oxford Poverty & Human Development Initiative (OPHI), Oxford Department of International Development (www.ophi.org.uk/wp-content/uploads/MPI_2011_Methodology_Note_4-11-2011_1500.pdf?79d835).

Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Yang. 2020. "Polarization and Public Health: Partisan Differences in Social Distancing During the Coronavirus Pandemic." *Journal of Public Economics* 191:104254 (https://doi.org/10.1016/j.jpubeco.2020.104254).

Allport, Gordon. 1954. *The Nature of Prejudice.* Cambridge, MA: Addison-Wesley.

Altheide, David L. and John M. Johnson. 1994. "Criteria for Assessing Interpretive Validity in Qualitative Research." Pp. 485-99 in *Handbook of Qualitative Research,* edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Altman, Lawrence K. 1998. "Getting It Right on the Facts of Death." *The New York Times,* December 22, D7.

Alwin, Duane F. and Jon A. Krosnick. 1991. "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes." *Sociological Methods & Research* 20:139-81.

Ambwani, Suman and Jaine Strauss. 2007. "Love Thyself Before Loving Others? A Qualitative and Quantitative Analysis of Gender Differences in Body Image and Romantic Love." *Sex Roles* 56:13-21.

American Association for Public Opinion Research (AAPOR). 2014. "Nonresponse in RDD Cell Phone Surveys" (www.aapor.org/AM/Template.cfm?Section=Cell_Phone_Task_Force_Report&Template=/CM/ContentDisplay.cfm&ContentID=3176#.U4JzqPldXTo).

American Medical Association (AMA). 2011. *AMA Code of Medical Ethics* (www.ama-assn.org/ama/physicianresources/medical-ethics/code-medicalethics.page?)

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Arlington, VA: American Psychiatric Publishing.

American Sociological Association (ASA). 1999. *Code of Ethics and Policies and Procedures of the ASA Committee on Professional Ethics.* Washington, DC: American Sociological Association.

American Sociological Association. 2018. *Code of Ethics.* (https://www.asanet.org/sites/default/files/asa_code_of_ethics-june2018a.pdf).

American Sociological Association (ASA). n.d. *Topic: Conflicts of Interest* (http://www.asanet.org/teaching-learning/faculty/teaching-ethics-throughout-curriculum/topic-conflicts-interest).

Anderson, Elijah. 1990. *Streetwise: Race, Class, and Change in an Urban Community.* Chicago, IL: University of Chicago Press.

Anderson, Elijah. 1999. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City.* New York, NY: Norton.

Anderson, Elijah. 2003. "Jelly's Place: An Ethnographic Memoir." *Symbolic Interaction.* 26:217-37.

Anderson, Lorin W. and David R. Krathwohl, eds. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* New York: Longman.

Anderson, Monica and Andrew W. Perrin. 2016. "13% of Americans Don't Use the Internet. Who Are They?" Pew Research Center, September 7 (http://www.pewresearch.org/fact-tank/2016/09/07/some-americans-dont-use-the-internetwho-are-they/).

Anderson, Rick. 2017. "Cabell's New Predatory Journal Blacklist: A Review." The Scholarly Kitchen, July 25 (https://scholarlykitchen.sspnet.org/2017/07/25/cabells-new-predatory-journal-blacklistreview/).

Anderton, Douglas L., Richard E. Barrett, and Donald J. Bogue. 1997. *The Population of the United States.* New York, NY: Free Press.

Andresen E., T. K. Malmstrom, F. D. Wolinsky, M. Schootman, J. P. Miller, and D. K. Miller. 2008. "Rating Neighborhoods for Older Adult Health: Results From the African American Health Study." *BMC Public Health* 8:35 (doi: 10.1186/1471-2458-8-35).

Aneshensel, Carol S. 2002. *Theory-Based Data Analysis for the Social Sciences.* Thousand Oaks, CA: Sage.

Anspach, Renee R. 1991. "Everyday Methods for Assessing Organizational Effectiveness." *Social Problems* 38(February):1-19.

Antaki, Charles. 2008. "Discourse Analysis and Conversation Analysis." Pp. 431-46 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Applebaum, Robert, Valerie Wellin, Cary Kart, J. Scott Brown, Heather Menne, Farida Ejaz, and Keren Brown Wilson. 2007. *Evaluation of Ohio's Assisted Living Medicaid Waiver Program: Final Summary Report.* Oxford, OH: Scripps Gerontology Center and Miami University.

Archibold, Randal C. 2010. "On Border Violence, Truth Pales Compared to Ideas." *The New York Times,* June 19 (www.nytimes.com).

Armas, Genaro C. 2002. "Government Won't File Appeal in Census Case." *The Boston Globe,* November 23:A1.

Armstrong, Karen. 2008. "Ethnography and Audience." Pp. 54-63 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Arthur, Michael W., John S. Briney, J. David Hawkins, Robert D. Abbott, Blair L. Brooke-Weiss, and Richard F. Catalano. 2007. "Measuring Risk and Protection in Communities Using the Communities That Care Youth Survey." *Evaluation and Program Planning* 30:197-211.

Arwood, Tracy and Sangeeta Panicker. 2007. "Assessing Risk in Social and Behavioral Sciences." *Collaborative Institutional Training Initiative* (https://www.citiprogram.org/members/learners).

Asch, Solomon E. 1958. "Interpersonal Influence." Pp. 174-83 in *Readings in Social Psychology,* 3rd ed., edited by Eleanor Maccoby, Theodore Newcomb, and Eugene Hartley. New York, NY: Holt, Rinehart & Winston.

Aseltine, Robert H. Jr. and Ronald C. Kessler. 1993. "Marital Disruption and Depression in a Community Sample." *Journal of Health and Social Behavior* 34(September):237-51.

Asher, Jeff, Ben Horwitz, and Toni Monkovic. 2021. "Why Does Louisiana Consistently Lead the Nation in Murders." *The New York Times,* February 16, A16.

The Associated Press. 2004. "'Get Tough' Youth Programs Are Ineffective, Panel Says." *The New York Times,* October 17, 25.

Ayhan, H. Öztaş. 2001. "Gender Bias in Agricultural Surveys Statistics by Gender: Measures to Reduce Gender Bias in Agricultural Surveys." *International Statistical Review* 69:447-60.

Babor, Thomas F., Robert S. Stephens, and G. Alan Marlatt. 1987. "Verbal Report Methods in Clinical Research on Alcoholism: Response Bias and Its Minimization." *Journal of Studies on Alcohol* 48(5):410-24.

Badger, Emily and Quoctrung Bui. 2020. "Pandemic Has Hindered Many of the Best Ideas for Reducing Gun Violence." *The New York Times,* A8.

Bagadiya, Jimit. 2017. "151 Amazing Social Media Statistics You Should Know in 2017." Social Pilot (https://www.socialpilot.co/blog/151-amazing-social-mediastatistics-know-2017).

Bail, Christopher A. 2008. "The Configuration of Symbolic Boundaries Against Immigrants in Europe." *American Sociological Review* 73:37-59.

Bainbridge, William Sims. 1989. *Survey Research: A Computer-Assisted Introduction.* Belmont, CA: Wadsworth.

Baker, Natalie D. and Magdalena Deham. 2020. "For a Short Time, We Were the Best Version of Ourselves: Hurricane Harvey and the Ideal of Community." *International Journal of Emergency Services* 9(1):8-20.

Ball, Richard A. and G. David. Curry. 1995. "The Logic of Definition in Criminology: Purposes and Methods for Defining 'Gangs.'" *Criminology* 33:225-45.

Banks, J. A. 1972. "Historical Sociology and the Study of Population." Pp. 55-70 in *Population and Social Change,* edited by D. V. Glass and Roger Revelle. London: Edward Arnold.

Bargh, John A., Katelyn Y. A. McKenna, and Grainne M. Fitzsimons. 2002. "Can You See Me? Activation and Expression of the 'True Self ' on the Internet." *Journal of Social Issues* 58:33-48.

Bargh, John A., Mark Chen, and Lara Burrows. 1996. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action." *Journal of Personality and Social Psychology* 71:230-44.

Barkan, Steven B. and Steven F. Cohn. 1994. "Racial Prejudice and Support for the Death Penalty by Whites." *Journal of Research in Crime and Delinquency* 31:202-09.

Barrera, Davide and Brent Simpson. 2012. "Much Ado About Deception: Consequences of Deceiving Research Participants in the Social Sciences." *Sociological Methods & Research* 41:383-413.

Barringer, Felicity. 1993. "Majority in Poll Back Ban on Handguns." *The New York Times,* June 4, A14.

Barry, John M. 2005. *The Great Influenza: The Story of the Deadliest Pandemic in History.* New York, NY: Penguin Books.

Bartholomew, David J. 2010. "Indirect Measurement." Pp. 455-68 in *The SAGE Handbook of Measurement,* edited by Geoffrey Walford, Eric Tucker, and Madhu Viswanathan. Thousand Oaks, CA: Sage.

Baum, Samuel. 1993. "Sources of Demographic Data." Pp. 3-1-3-50 in *Readings in Population Research Methodology. Vol. 1, Basic Tools,* edited by Donald J. Bogue, Eduardo E. Arriaga, and Douglas L. Anderton. Chicago, IL: Social Development Center, for the United Nations Population Fund.

Baumann, Jeannie. 2017. "Proposed Delay to Research Rule Still Up in the Air." From Health Care on Bloomberg Law, October 18 (https://www.bna.com/proposed-delay-research-n73014471089/).

Baumer, Eric P. 2002. "Neighbourhood Disadvantage and Police Notification by Victims of Violence." *Criminology* 40(3):579-616.

Baumrind, Diana. 1964. "Some Thoughts on Ethics of Research: After Reading Milgram's 'Behavioral Study of Obedience.'" *American Psychologist* 19:421-23.

Baumrind, Diana. 1985. "Research Using Intentional Deception: Ethical Issues Revisited." *American Psychologist* 40:165-74.

Bean, Lee L., Geraldine P. Mineau, and Douglas L. Anderton. 1990. *Fertility Change on the American Frontier: Adaptation and Innovation.* Berkeley, CA: University of California Press.

Becker, Howard S. 1958. "Problems of Inference and Proof in Participant Observation." *American Sociological Review* 23:652-60.

Becker, Howard S. 1963. *The Outsiders: Studies in the Sociology of Deviance.* New York, NY: Free Press.

Becker, Howard S. 1974. "Photography and Sociology." *Studies in the Anthropology of Visual Communication* 1:3-26.

Becker, Howard S. 2007. *Writing for Social Scientists,* 2nd ed. Chicago, IL: University of Chicago Press.

Bellah, Robert N., Richard Madsen, William M. Sullivan, Ann Swidler, and Steven M. Tipton. 1985. *Habits of the Heart: Individualism and Commitment in American Life.* New York, NY: Harper & Row.

Belousov, Konstantin, Tom Horlick-Jones, Michael Bloor, Yakov Gilinskiy, Valentin Golbert, Yakov Kostikovsky, Michael Levi, and Dmitri Pentsov. 2007. "Any Port in a Storm: Fieldwork Difficulties in Dangerous and Crisis-Ridden Settings." *Qualitative Research* 7:155-75.

Bendix, Reinhard. 1956. *Work and Authority in Industry: Ideologies of Management in the Course of Industrialization.* Berkeley, CA: University of California Press.

Bendix, Reinhard. 1962. *Max Weber: An Intellectual Portrait.* Garden City, NY: Doubleday/Anchor.

Benedek, Thomas G. 1978. "The 'Tuskegee Study' of Syphilis: Analysis of Moral Versus Methodologic Aspects." *Journal of Chronic Diseases* 31(1):35-50.

Benedek, Thomas G. and Jonathon Erlen. 1999. "The Scientific Environment of the Tuskegee Study of Syphilis, 1920-1960." *Perspectives in Biology and Medicine* 43(1):1-30.

Benkler, Yochai. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom.* New Haven, CT: Yale University Press.

Benwell, Bethan and Elizabeth Stokoe. 2006. *Discourse and Identity.* Edinburgh, Scotland: Edinburgh University Press.

Berger, Joseph, Bernard P. Cohen, and Morris Zelditch Jr. 1972. "Status Characteristics and Social Interaction." *American Sociological Review* 37:241-55.

Bergman, Manfred Max. 2008. "Combining Different Types of Data for Quantitative Analysis." Pp. 585–601 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Berk, Richard A., Alec Campbell, Ruth Klap, and Bruce Western. 1992. "The Deterrent Effect of Arrest: A Bayesian Analysis of Four Field Experiments." *American Sociological Review* 57 (October):698-708.

Berk, Richard A., Kenneth J. Lenihan, and Peter H. Rossi. 1980. "Crime and Poverty: Some Experimental Evidence From Ex-offenders." *American Sociological Review* 45:766-86.

Berkman, Lisa F. and S. Leonard Syme. 1979. "Social Networks, Host Resistance, and Mortality: A Nine-Year Follow-Up Study of Alameda County Residents." *American Journal of Epidemiology* 109(2):186-204.

Berman, Greg and Aubrey Fox. 2009. *Lessons From the Battle Over D.A.R.E.: The Complicated Relationship Between Research and Practice.* Washington, DC: Center for Court Innovation and Bureau of Justice Assistance, U.S. Department of Justice.

Berry, Brent. 2006. "Friends for Better or for Worse: Interracial Friendship in the United States as Seen Through Wedding Party Photos." *Demography* 43:491-510.

Binder, Arnold and James W. Meeker. 1993. "Implications of the Failure to Replicate the Minneapolis Experimental Findings." *American Sociological Review* 58(December):886-88.

Bingham, Andrea J. and Patricia Witkowsky. 2022. "Deductive and Inductive Approaches to Qualitative Analysis." Pp. 97-148 in *Analyzing and Interpreting Qualitative Research: After the Interview,* edited by C. Vanover, P. Mihas, and J. Saldana. Thousand Oaks, CA: Sage.

Birkeland, Sarah, Erin Murphy-Graham, and Carol Weiss. 2005. "Good Reasons for Ignoring Good Evaluation: The Case of the Drug Abuse Resistance Education (D.A.R.E.) Program." *Evaluation and Program Planning* 28:247-56.

Black, Donald J., ed. 1984. *Toward a General Theory of Social Control.* Orlando, FL: Academic Press.

Black, M. C., K. C. Basile, M. J. Breiding, S. G. Smith, M. L. Walters, M. T. Merrick, J. Chen, & M. R. Stevens. 2011. *The National Intimate Partner and Sexual Violence Survey (NISVS): 2010 Summary Report.* Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Blanco, Gerardo L. and Gretchen B. Rossman. 2022. "As a Qualitative Study Unfolds: Shifts in Design and Analysis." Pp. 7-22 in *Analyzing and Interpreting Qualitative Research: After the Interview,* edited by C. Vanover, P. Mihas, and J. Saldana. Thousand Oaks, CA: Sage.

Blau, Peter M. 1964. *Exchange and Power in Social Life.* New York, NY: Wiley.

Bledsoe, Katrina L. and Rodney K. Hopson. 2009. "Conducting Ethical Research and Evaluation in Underserved Communities." Pp. 391-406 in *The Handbook of Social Research Ethics,* edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Bloom, Howard S. 2008. "The Core Analytics of Randomized Experiments for Social Research." Pp. 115-33 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Blow, Charles M. 2021. "A Holistic View of Vaccine Hesitancy." *The New York Times,* February 7.

Boase, Jeffrey, John B. Horrigan, Barry Wellman, and Lee Rainie. 2006. *The Strength of Internet Ties: The Internet and Email Aid Users in Maintaining Their Social Networks and Provide Pathways to Help When People Face Big Decisions.* Washington, DC: Pew Internet & American Life Project.

Bogdewic, Stephan P. 1999. "Participant Observation." Pp. 47-70 in *Doing Qualitative Research,* 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Bogue, Donald J. 1969. *Principles of Demography.* New York, NY: Wiley.

Bogue, Donald J., Eduardo E. Arriaga, and Douglas L. Anderton. 1993. *Readings in Population Research Methodology. Vol. 1, Basic Tools.* Chicago, IL: Social Development Center, for the United Nations Population Fund.

Bohannon, John. 2013. "Who's Afraid of Peer Review?" *Science* 6154:60-5.

Bohn, Angela, Christian Buchta, Kurt Hornik, and Patrick Mair. 2014. "Making Friends and Communicating on Facebook: Implications for the Access to Social Capital." *Social Networks* 37:29-41.

Bollen, Kenneth A., Barbara Entwisle, and Arthur S. Alderson. 1993. "Macrocomparative Research Methods." *Annual Review of Sociology* 19:321-51.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489:295-98.

Booth, Andrew, Anthea Sutton, and Diana Papaioannou. 2016. *Systematic Approaches to a Successful Literature Review,* 2nd ed. Thousand Oaks, CA: Sage.

Booth, Brenda M., Greer Sullivan, Paul Koegel, and Audrey Burnam. 2002. "Vulnerability Factors for Homelessness Associated With Substance Dependence in a Community Sample of Homeless Adults." *American Journal of Drug and Alcohol Abuse* 28:429-52.

Booth, Wayne C., Gregory G. Colomb, and Joseph M. Williams. 1995. *The Craft of Research.* Chicago, IL: University of Chicago Press.

Booth, Wayne C., Gregory G. Colomb, and Joseph M. Williams. 2008. *The Craft of Research,* 3rd ed. Chicago, IL: University of Chicago Press.

Borg, Marian J. 1997. "The Southern Subculture of Punitiveness? Regional Variation in Support for Capital Punishment." *Journal of Research in Crime and Delinquency* 34:25-45.

Borg, Marian J. 1998. "Vicarious Homicide Victimization and Support for Capital Punishment: A Test of Black's Theory of Law." *Criminology* 36:537-67.

Boruch, Robert F. 1997. *Randomized Experiments for Planning and Evaluation: A Practical Guide.* Thousand Oaks, CA: Sage

Bos, Johannes, Aletha Huston, Robert Granger, Greg Duncan, Thomas Brock, and Vonnie McLoyd. 1999. *New Hope for People With Low Incomes: Two-Year Results of a Program to Reduce Poverty and Reform Welfare.* MRDC Report (www.mdrc.org/publication/new-hope-people-low-incomes).

Bosk, Charles L. and Raymond G. De Vries. 2004. "Bureaucracies of Mass Deception: Institutional Review Boards and the Ethics of Ethnographic Research." *Annals of the American Academy of Political and Social Science* 595:249-63.

Bosman, Julie, Jan Hoffman, and Margot Sanger-Katz. 2021. "Who Are the Unvaccinated in America?" *The New York Times,* August 1.

Bourgois, Philippe, Mark Lettiere, and James Quesada. 1997. "Social Misery and the Sanctions of Substance Abuse: Confronting HIV Risk Among Homeless Heroin Addicts in San Francisco." *Social Problems* 44:155-73.

Braga, Anthony A., Lisa M. Barao, Gregory M. Zimmerman, Stephen Douglas, and Keller Sheppard. 2020. "Measuring the Direct and Spillover Effects of Body Worn Cameras on the Civility of Police–Citizen Encounters and Police Work Activities." *Journal of Quantitative Criminology* 36(4):851-76.

Braga, Anthony A., William H. Sousa, James R. Coldren, and Denise Rodriguez. 2018. "The Effects of Body-Worn Cameras on Police Activity and Police-Citizen Encounters: A Randomized Controlled Trial." *Journal of Criminal Law and Criminology* 108(3):511-38.

Branquinho, Cátia, Colette Kelly, Lourdes C. Arevalo, Anabela Santos, and Margarida Gaspar de Matos. 2020. "'Hey, We Also Have Something to Say': A Qualitative Study of Portuguese Adolescents' and Young People's Experiences Under COVID-19." *Journal of Community Psychology* 48(8):2740-52.

Bray, Hiawatha. 1999. "Plugging in to the Electronic Campus." *The Boston Globe Magazine,* April 11, 20-30.

Brenner, Philip S. 2012. "Investigating the Effect of Bias in Survey Measures of Church Attendance." *Sociology of Religion* 73:361-83.

Brenner, Philip S. 2021. "Effects of Nonresponse, Measurement, and Coverage Bias in Survey Estimates of Voting." *Social Science Quarterly* 102(2):939-54.

Brewer, John and Albert Hunter. 1989. *Multi-method Research: A Synthesis of Styles.* New-bury Park, CA: Sage.

Bridges, George S. and Joseph G. Weis. 1989. "Measuring Violent Behavior: Effects of Study Design on Reported Correlates of Violence." Pp. 14-34 in *Violent Crime, Violent Criminals,* edited by Neil Alan Weiner and Marvin E. Wolfgang. Newbury Park, CA: Sage.

Brooks, Clem and Jeff Manza. 2006. "Social Policy Responsiveness in Developed Democracies." *American Sociological Review* 71:474-94.

Broskoske, Steve. 2005. "How to Prevent Paper Recycling." *The Teaching Professor* 19:1, 4.

Brown, Bruce L. and Dawson Hedges. 2009. "Use and Misuse of Quantitative Methods: Data Collection, Calculation, and Presentation." Pp. 373-86 in *The Handbook of Social Research Ethics,* edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Brown, Judith Belle. 1999. "The Use of Focus Groups in Clinical Research." Pp. 109-24 in *Doing Qualitative Research,* 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Brown, Ryan A., David P. Kennedy, Joan S. Tucker, Daniela Golinelli, and Suzanne L. Wenzel. 2013. "Monogamy on the Street: A Mixed Methods Study of Homeless Men." *Journal of Mixed Methods Research* 7:328-46.

Bruni, Frank. 2002. "Persistent Drop in Fertility Reshapes Europe's Future." *The New York Times,* December 26, A1, A10.

Bureau of Justice Statistics. 2011. *Key Facts at a Glance: Drug Arrests by Age, 1970–2007.* Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice (http://bjs.ojp.usdoj.gov/content/glance/drug.cfm).

Burke, Garance. 2000. "Mexico's Census Battles Perception of Corruption." *The Boston Globe,* March 2, A2.

Burt, Martha R. 1996. "Homelessness: Definitions and Counts." Pp. 15-23 in *Homelessness in America,* edited by Jim Baumohl. Phoenix, AZ: Oryx.

Busblog. 2005. "Yesterday I Had to Go to Jury Duty." Retrieved August 31, 2021 (http://busblog.tonypierce.com/2005/06/yesterday-i-had-to-go-to-jury-duty-to.html).

Bushman, Brad J., Roy F. Baumeister, and Angela D. Stack. 1999. "Catharsis, Aggression, and Persuasive Influence: Self-Fulfilling or Self-Defeating Prophecies?" *Journal of Personality and Social Psychology* 76:367-76.

Butler, Declan. 2013. "When Google Got Flu Wrong." *Nature* 494(February 14):155-56.

Butler, Dore and Florence Geis. 1990. "Nonverbal Affect Responses to Male and Female Leaders: Implications for Leadership Evaluations." *Journal of Personality and Social Psychology* 58(January):48-59.

Buttel, Frederick H. 2000. "World Society, the Nation-State, and Environmental Protection: Comment on Frank, Hironaka, and Schofer." *American Sociological Review* 65:117-21.

Buzawa, Eve S. and Carl G. Buzawa (Eds.). 1996. *Do Arrests and Restraining Orders Work?* Thousand Oaks, CA: Sage.

Cain, Carol. 1991. "Personal Stories: Identity Acquisition and Self-Understanding in Alcoholics Anonymous." *Ethos* 19:210-53.

Callaghan, Timothy, Ali Moghtaderi, Jennifer A. Lueck, Peter Hotez, Ulrich Strych, Avi Dor, Erika Franklin Fowler, and Matthew Motta. 2021. "Correlates and Disparities of Intention to Vaccinate Against COVID-19." *Social Science and Medicine* 272.

Calvey, David. 2017. *Covert Research: The Art, Politics and Ethics of Undercover Fieldwork.* Thousand Oaks, CA: Sage.

Campbell, Colin and Jonathan Horowitz. 2016. "Does College Influence Sociopolitical Attitudes?" *Sociology of Education* 89(1):40-58.

Campbell, Donald T. and Donald W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait–Multimethod Matrix." *Psychological Bulletin* 56:81-105.

Campbell, Donald T. and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research.* Chicago, IL: Rand McNally.

Campbell, Donald T. and M. Jean Russo. 1999. *Social Experimentation.* Thousand Oaks, CA: Sage.

Campbell, Richard T. 1992. "Longitudinal Research." Pp. 1146-58 in *Encyclopedia of Sociology,* edited by Edgar F. Borgatta and Marie L. Borgatta. New York, NY: Macmillan.

Campbell, Wilson. 2002. "A Statement From the Governmental Accounting Standards Board and Performance Measurement Staff." American Society for Public Administration (www.aspanet.org/cap/forum_statement.html#top).

Cannell, C. F., L. Oksenberg, G. Kalton, K. Bischoping, and Floyd J. Fowler. 1989. *New Techniques for Pretesting Survey Questions.* Final Report. Grant No. HS05616. National Center for Health Service Research, Health Care Technology Assessment (www.psc.isr.umich.edu/dis/infoserv/isrpub/pdf/NewTechniquesPretestingSurveyQuestions_OCR.pdf).

Caplan, Arthur L. 1992. "Twenty Years after: The Legacy of the Tuskegee Syphilis Study." *Hastings Center Report* 22:32-5.

Caplan, Brina, Russell K. Schutt, Winston M. Turner, Stephen M. Goldfinger, and Larry J. Seidman. 2006. "Change in Neurocognition by Housing Type and Substance Abuse Among Formerly Homeless Seriously Mentally Ill Individuals." *Schizophrenia Research* 83:77-86.

Caputi, Theodore L. and Thomas McLellan. 2017. "Truth and D.A.R.E.: Is D.A.R.E.'s Keepin' it REAL Suitable for American Nationwide Implementation?" *Drugs: Education, Prevention and Policy* 24(1):49-57.

Carmines, Edward G. and Richard A. Zeller. 1979. *Reliability and Validity Assessment,* no. 17, *Quantitative Applications in the Social Sciences.* Beverly Hills, CA: Sage.

Carney, Megan A. 2017. "'Back There We Had Nothing to Eat': The Case of Transnational Food Insecurity." *International Migration* 55(4):64-77.

Casanova, José Luís. 2018. "Crisis and Cultural Change: The Countries With Adjustment Programmes in the European Union." *Comparative Sociology* 17:187-220.

Casselman, Ben and Ella Koeze. 2021. "The Pandemic Changed How We Spent Our Time." *The New York Times,* July 29, B1, B6.

Casselman, Ben and Jeanna Smialek. 2021. "U.S. Poverty Fell Last Year as Government Aid Made Up for Lost Jobs." *New York Times,* September 14.

Catalano, Shannan. 2012. *Intimate Partner Violence, 1993–2010.* Washington, DC: Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

Cava, Anita, Reid Cushman, and Kenneth Goodman. 2007. "HIPAA and Human Subjects Research." *Collaborative Institutional Training Initiative* (https://www.citiprogram.org/members/learners).

Cave, Emma and Soren Holm. 2003. "Milgram and Tuskegee—Paradigm Research Projects in Bioethics." *Health Care Analysis* 11:27-40.

Ceglowski, Deborah. 2002. "Research as Relationship." Pp. 5-27 in *The Qualitative Inquiry Reader,* edited by Norman Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Center for Survey Research, University of Massachusetts at Boston. 1987. "Methodology: Designing Good Survey Questions." *Newsletter,* April, 3.

Center for Survey Research. 2021. *Living in Boston During COVID-19 Survey.* Boston, MA: University of Massachusetts Boston.

Centers for Disease Control and Prevention (CDC). 2009. *The Tuskegee Timeline.* Atlanta, GA: National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention (www.cdc.gov/tuskegee/timeline.htm).

Chadwick, Gary L. 2017. "Final Rule Material: New and Revised Definitions." CITI Program (https://about.citiprogram.org/wp-content/uploads/2017/07/Final-Rule-Material-New-and-Revised-Definitions.pdf).

Chase-Dunn, Christopher and Thomas D. Hall. 1993. "Comparing World-Systems: Concepts and Working Hypotheses." *Social Forces* 71:851-86.

Chen, Huey-Tsyh. 1990. *Theory-Driven Evaluations.* Newbury Park, CA: Sage.

Chen, Huey-Tsyh and Peter H. Rossi. 1987. "The Theory-Driven Approach to Validity." *Evaluation and Program Planning* 10:95-103.

Christian, Leah, Scott Keeter, Kristen Purcell, and Aaron Smith. 2010. *Assessing the Cell Phone Challenge to Survey Research in 2010.* Washington, DC: Pew Research Center for the People & the Press and Pew Internet & American Life Project.

Church, A. Timothy. 2010. "Measurement Issues in Cross-Cultural Research." Pp. 151-75 in *The SAGE Handbook of Measurement,* edited by Geoffrey Walford, Eric Tucker, and Madhu Viswanathan. Thousand Oaks, CA: Sage.

Church, Allan H. 1993. "Estimating the Effects of Incentives on Mail Survey Response Rates: A Meta-Analysis." *Public Opinion Quarterly* 57:62-79.

CIRCLE. 2013. *Fact Sheet: Voter Turnout Among Young Women and Men in the 2012 Presidential Election.* Medford, MA: The Center for Information & Research on Civil Learning & Engagement, Jonathan M. Tisch College of Citizenship and Public Service, Tufts University (www.civicyouth.org/wp-content/uploads/2013/05/fs_gender_13_final.pdf).

CITI Program. 2017. *Final Rule Resources* (https://about.citiprogram.org/en/final-rule-resources/).

City of Philadelphia and Institute for Survey Research. 2021. Philadelphia Resident Survey Report, 2019-2020. Philadelphia: City of Philadelphia and Institute for Survey Research, Temple University. 2019ResidentSurvey-FINAL.pdf (phila.gov)

Clare, Joseph, Darren Henstock, Christine McComb, Roy Newland, and Geoffrey C. Barnes. 2021. "The Results of a Randomized Controlled Trial of Police Body-Worn Video in Australia." *Journal of Experimental Criminology* 17(1):43-54.

Clark, Herbert H. and Michael F. Schober. 1994. "Asking Questions and Influencing Answers." Pp. 15-48 in *Questions About Questions: Inquiries Into the Cognitive Bases of Surveys,* edited by Judith M. Tanur. New York, NY: Russell Sage Foundation.

Clement, Scott and Peyton M. Craighill. 2015. Gallup Isn't Doing Any Horserace Polling in 2016. Here's Why. *The Washington Post,* October 7 (https://www.washingtonpost.com/news/the-fix/wp/2015/10/07/gallupisnt-doing-any-horserace-polling-in-2016-heres-why/?utm_term=.e1cae30e7250).

Clinton, William Jefferson. 1997. *Remarks by the President in Apology for Study Done in Tuskegee.* Washington, DC: Office of the Press Secretary, The White House (https://clintonwhitehouse4.archives.gov/textonly/New/Remarks/Fri/19970516-898.html).

Coffey, Amanda and Paul Atkinson. 1996. *Making Sense of Qualitative Data: Complementary Research Strategies.* Thousand Oaks, CA: Sage.

Cohen, Alison K. and Benjamin W. Chaffee. 2012. "The Relationship Between Adolescents' Civic Knowledge, Civic Attitude, and Civic Behavior and Their Self-Reported Future Likelihood of Voting." *Education, Citizenship and Social Justice* 8(1):43-57.

Cohen, Elizabeth and Dana Vigue. 2020. "Covid-19 Vaccine Trials Have Been Slow to Recruit Black and Latino People—and That Could Delay a Vaccine - CNN." August 16.

Cohen, Sheldon, Robin Mermelstein, Tom Kamarck, and Harry M. Hoberman. 1985. "Measuring the Functional Components of Social Support." Pp. 73-94 in *Social Support: Theory, Research and Applications,* edited by Irwin G. Sarason and Barbara R. Sarason. The Hague, The Netherlands: Martinus Nijhoff.

Cohen, Susan G. and Gerald E. Ledford Jr. 1994. "The Effectiveness of Self-Managing Teams: A Quasi-Experiment." *Human Relations* 47:13-43.

Cohn, Nate. 2017. "Election Review: Why Crucial State Polls Turned Out to Be Wrong." *The New York Times,* June 1:A12.

Coleman, James S. 1990. *Foundations of Social Theory.* Cambridge, MA: Harvard University Press.

Coleman, James S. and Thomas Hoffer. 1987. *Public and Private High Schools: The Impact of Communities.* New York, NY: Basic Books.

Coleman, James S., Thomas Hoffer, and Sally Kilgore. 1982. *High School Achievement: Public, Catholic, and Private Schools Compared.* New York, NY: Basic Books.

Coleman-Minahan, Kate and Goleen Samari. 2020. "'He Supported Me 100%': Mexican-Immigrant Fathers, Daughters, and Adolescent Sexual Health." *Ethnicity and Health* 25(4):560-79.

College Alcohol Study. 2008. "College Alcohol Study News." *Harvard School of Public Health* (www.hsph.harvard.edu/cas/Home.html).

Collins, Randall. 1994. *Four Sociological Traditions.* New York: Oxford University Press.

Connolly, Francis J. and Charley Manning. 2001. "What 'Push Polling' Is and What It Isn't." *The Boston Globe,* August 16, A21.

Connor, Jennie, Andrew Gray, and Kypros Kypri. 2010. "Drinking History, Current Drinking and Problematic Sexual Experiences Among University Students." *Australian and New Zealand Journal of Public Health* 34:487-94.

Connors, Gerard J. and Robert J. Volk. 2004. "Self-Report Screening for Alcohol Problems Among Adults." Pp. 21-35 in *Assessing Alcohol Problems: A Guide for Clinicians and Researchers,* 2nd ed., edited by John P. Allen and Veronica B. Wilson. NIH Publication No. 03 -3745, revised 2003. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health (Retrieved October 6, 2005, from https://pubs.niaaa.nih.gov/publications/AssessingAlcohol/index.htm).

Converse, Jean M. 1984. "Attitude Measurement in Psychology and Sociology: The Early Years." Pp. 3-40 in *Surveying Subjective Phenomena,* vol. 2, edited by Charles F. Turner and Elizabeth Martin. New York, NY: Russell Sage Foundation.

Cook, Gareth. 2005. "Face It: Appearances Sway Voters." *The Boston Globe,* June 14, E1, E4.

Cook, Thomas D. and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Chicago, IL: Rand McNally.

Cook, Thomas D. and Vivian C. Wong. 2008. "Better Quasi-Experimental Practice." Pp. 134-65 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Cooper, Harris and Larry V. Hedges. 1994. "Research Synthesis as a Scientific Enterprise." Pp. 3-14 in *The Handbook of Research Synthesis,* edited by Harris Cooper and Larry V. Hedges. New York, NY: Russell Sage Foundation.

Cooper, Kathleen B. and Nancy J. Victory. 2002. "Foreword." P. iv in *A Nation Online: How Americans Are Expanding Their Use of the Internet.* Washington, DC: National Telecommunications and Information Administration, Economics and Statistics Administration, U.S. Department of Commerce.

Core Institute. 1994. *Core Alcohol and Drug Survey: Long Form.* Carbondale, IL: FIPSE Core Analysis Grantee Group, Core Institute, Student Health Programs, Southern Illinois University.

Correll, Joshua, Bernadette Park, Charles M. Judd, and Bernd Wittenbrink. 2002. "The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals." *Journal of Personality and Social Psychology* 83:1314-29.

Corse, Sara J., Nancy B. Hirschinger, and David Zanis. 1995. "The Use of the Addiction Severity Index With People With Severe Mental Illness." *Psychiatric Rehabilitation Journal* 19(1):9-18.

Costner, Herbert L. 1989. "The Validity of Conclusions in Evaluation Research: A Further Development of Chen and Rossi's Theory-Driven Approach." *Evaluation and Program Planning* 12:345-53.

Couper, Mick P. 2000. "Web Surveys: A Review of Issues and Approaches." *Public Opinion Quarterly* 64:464-94.

Couper, Mick P. and Peter V. Miller. 2008. "Web Survey Methods: Introduction." *Public Opinion Quarterly* 72:831-35.

Couper, Mick P., Michael W. Traugott, and Mark J. Lamias. 2001. "Web Survey Design and Administration." *Public Opinion Quarterly* 65:230-53.

Couper, Mick P., Reginald P. Baker, Jelke Bethlehem, Cynthia Z. F. Clark, Jean Martin, William L. Nicholls II, and James M. O'Reilly (Eds.). 1998. *Computer-Assisted Survey Information Collection.* New York, NY: Wiley.

Cowan, Gloria. 2002. "Content Analysis of Visual Materials." Pp. 345-68 in *Handbook for Conducting Research on Human Sexuality,* edited by Michael W. Wiederman and Bernard E. Whitley Jr. London: Lawrence Erlbaum.

Cress, Daniel M. and David A. Snow. 2000. "The Outcomes of Homeless Mobilization: The Influence of Organization, Disruption, Political Mediation, and Framing." *American Journal of Sociology* 4:1063-104.

Creswell, John W. and Vicki L. Plano Clark. 2011. *Designing and Conducting Mixed Methods Research,* 2nd ed. Thousand Oaks, CA: Sage.

Cuba, Lee J. 2002. *A Short Guide to Writing About Social Science,* 4th ed. New York, NY: Addison-Wesley.

Cullen, Kevin. 2014. "The Gunmen, the Shadows, and the Damage Done." *Boston Sunday Globe,* July 6, A1, A6, A7.

Currie, Janet and Duncan Thomas. 1995. "Does Head Start Make a Difference?" *American Economic Review* 85:341-65.

Czopp, Alexander M., Margo J. Monteith, and Aimee Y. Mark. 2006. "Standing Up for a Change: Reducing Bias Through Interpersonal Confrontation." *Journal of Personality and Social Psychology* 90:784-803.

D.A.R.E. 2008. *The "New" D.A.R.E. Program* Retrieved from (www.dare.com/newdare.asp).

D.A.R.E. 2014. "Drug Abuse Resistance Education" (www.dare.org/starting-a-dare-program).

D'Amico, Elizabeth J. and Kim Fromme. 2002. "Brief Prevention for Adolescent Risk-Taking Behavior." *Addiction* 97:563–574.

Dabney, Dean A., Laura Dugan, Volkan Topalli, and Richard C. Hollinger. 2006. "The Impact of Implicit Stereotyping on Offender Profiling: Unexpected Results From an Observational Study of Shoplifting." *Criminal Justice and Behavior* 33:646-75.

Dahlberg, Britt, Marsha N. Wittink, and Joseph J. Gallo. 2010. "Funding and Publishing Integrated Studies: Writing Effective Mixed Methods Manuscripts and Grant Proposals." Pp. 775-802 in *SAGE Handbook of Mixed Methods in Social & Behavioral Research,* 2nd ed., edited by Abbas Tashakkori and Charles Teddlie. Thousand Oaks, CA: Sage.

Dale, Angela, Jo Wathan, and Vanessa Higgins. 2008. "Secondary Analysis of Quantitative Data Sources." Pp. 520-35 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Dannefer, W. Dale and Russell K. Schutt. 1982. "Race and Juvenile Justice Processing in Court and Police Agencies." *American Journal of Sociology* 87(March):1113-32.

Davern, Michael, René Bautista, Jeremy Freese, Stephen L. Morgan, and Tom W. Smith. 2021. *General Social Survey 2021 Cross-Section.* [Machine-Readable Data File]. Chicago, IL: NORC.

Davis, James A. 1985. *The Logic of Causal Order.* Sage University Paper Series on Quantitative Applications in the Social Sciences, series No. 07–055. Beverly Hills, CA: Sage.

Davis, James A. and Tom W. Smith. 1992. *The NORC General Social Survey: A User's Guide.* Newbury Park, CA: Sage.

Davis, Ophera A. and Marie Land. 2007. "Southern Women Survivors Speak About Hurricane Katrina, the Children and What Needs to Happen Next." *Race, Gender & Class* 14:69-86.

Davis, R. C., D. Weisburd, and B. Taylor. 2008. "Effects of Second Responder Programs on Repeat Incidents of Family Abuse." *Campbell Systematic Reviews,* 15.

Dawes, Robyn. 1995. "How Do You Formulate a Testable Exciting Hypothesis?" Pp. 93-6 in *How to Write a Successful Research Grant Application: A Guide for Social and Behavioral Scientists,* edited by Willo Pequegnat and Ellen Stover. New York, NY: Plenum Press.

de Leeuw, Edith. 2008. "Self-Administered Questionnaires and Standardized Interviews." Pp. 311-27 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

de Vaus, David. 2008. "Comparative and Cross-National Designs." Pp. 248–264 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

De, Rahul, Neena Pandey, and Abhipsa Pal. 2020. "Since January 2020 Elsevier Has Created a COVID-19 Resource Centre With Free Information in English and Mandarin on the Novel Coronavirus COVID-19. The COVID-19 Resource Centre Is Hosted on Elsevier Connect, the Company 's Public News and Information." *International Journal of Information and Learning Technology* (55):1-6.

Decker, Scott H. and Barrik Van Winkle. 1996. *Life in the Gang: Family, Friends, and Violence.* Cambridge: Cambridge University Press.

Deegan, Allison. 2012. "Stranger in a Strange Land: The Challenges and Benefits of Online Interviews in the Social Networking Space." Pp. 69-99 in *Cases in Online Interview Research,* edited by Janet Salmons. Thousand Oaks, CA: Sage.

Demos, John. 1998. "History Beyond Data Bits." *The New York Times,* December 30, A23.

Dentler, Robert A. 2002. *Practicing Sociology: Selected Fields.* Westport, CT: Praeger.

Denzin, Norman K. 2002. "The Interpretive Process." Pp. 349-68 in *The Qualitative Researcher's Companion,* edited by A. Michael Huberman and Matthew B. Miles. Thousand Oaks, CA: Sage.

Denzin, Norman and Yvonna S. Lincoln. 2000. "Introduction: The Discipline and Practice of Qualitative Research." Pp. 1-28 in *The Handbook of Qualitative Research,* 2nd ed., edited by Norman Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Denzin, Norman K. and Yvonna S. Lincoln. 2008. *Strategies of Qualitative Inquiry,* 3rd ed. Thousand Oaks, CA: Sage.

DeParle, Jason. 1999. "Project to Rescue Needy Stumbles Against the Persistence of Poverty." *The New York Times,* May 15, A1, A10.

Department of Health and Human Services (DHHS). 2009. Code of Federal Regulations, Title 45: Public Welfare. Department of Health and Human Services, Part 46, Protection of Human Subjects (hhs.gov/ohrp/policy/ohrpregulations.pdf).

Department of Health, Education, and Welfare. 1979. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research.* Washington, DC: The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Office of the Secretary, Department of Health, Education, and Welfare (www.hss.gov/ohrp/humansubjects/guidance/belmont.htm).

DeSilver, Drew. 2016. "Turnout Was High in the 2016 Primary Season, but Just Short of 2008 Record." Pew Research Center, June 10 (http://www.pewresearch.org/fact-tank/2016/06/10/turnout-was-high-inthe-2016-primary-season-but-just-shortof-2008-record/).

DeSilver, Drew. 2021. *Turnout Soared in 2020 as Nearly Two-thirds of Eligible U.S. Voters Cast Ballots for President.* Washington, DC: Pew Research Center, January 28.

Desjardins, Jeff. 2017. "What Happens in an Internet Minute in 2017?" (http://www.visualcapitalist.com/happens-internet-minute-2017/).

Desmond, Matthew, Andrew Papachristos, and David Kirk. 2016. "Police Violence and Citizen Crime Reporting in the Black Community." *American Sociological Review* 81(5):857-76.

DeVellis, Robert F. 2017. *Scale Development: Theory and Applications,* 4th ed. Thousand Oaks, CA: Sage.

Dewan, Shaila K. 2004a. "As Murders Fall, New Tactics Are Tried Against Remainder." *The New York Times,* December 31, A24-A25.

Dewan, Shaila K. 2004b. "New York's Gospel of Policing by Data Spreads Across U.S." *The New York Times,* April 26, A1, C16.

di Gregorio, Silvana. 2022. "Voice to Text: Automating Transcription." Pp. 97–116 in *Analyzing and Interpreting Qualitative Research: After the Interview,* edited by C. Vanover, P. Mihas, and J. Saldana. Thousand Oaks, CA: Sage..

Diamond, Timothy. 1992. *Making Gray Gold: Narratives of Nursing Home Care.* Chicago, IL: University of Chicago Press.

Dickson-Swift, Virginia, Erica L. James, Sandra Kippen, and Pranee Liamputtong. 2008. "Risk to Researchers in Qualitative Research on Sensitive Topics: Issues and Strategies." *Qualitative Health Research* 18:133-44.

DiClemente, C. C., J. P. Carbonari, R. P. G. Montgomery, and S. O. Hughes. 1994. "The Alcohol Abstinence Self-Efficacy Scale." *Journal of Studies on Alcohol* 55:141-48.

Dill, Karen E. and Kathryn P. Thill. 2007. "Video Game Characters and the Socialization of Gender Roles: Young People's Perceptions Mirror Sexist Media Depictions." *Sex Roles* 57:851-64.

Dillman, Don A. 1978. *Mail and Telephone Surveys: The Total Design Method.* New York, NY: Wiley.

Dillman, Don A. 1982. "Mail and Other Self-Administered Questionnaires." Chapter 12 in *Handbook of Survey Research,* edited by Peter Rossi, James Wright, and Andy Anderson. New York: Academic Press. As reprinted on pp. 637-38 in Delbert C. Miller, 1991. *Handbook of Research Design and Social Measurement,* 5th ed. Beverly Hills, CA: Sage.

Dillman, Don A. 2000. *Mail and Internet Surveys: The Tailored Design Method,* 2nd ed. New York, NY: Wiley.

Dillman, Don A. 2007. *Mail and Internet Surveys: The Tailored Design Method,* 2nd ed. Updated With New Internet, Visual, and Mixed-Mode Guide. Hoboken, NJ: Wiley.

Dillman, Don A. and Leah Melani Christian. 2005. "Survey Mode as a Source of Instability in Responses Across Surveys." *Field Methods* 17:30-52.

Dillman, Don A., James A. Christenson, Edwin H. Carpenter, and Ralph M. Brooks. 1974. "Increasing Mail Questionnaire Response: A Four-State Comparison." *American Sociological Review* 39(October):744-56.

Donais, Lauren, Bill Simonsen, and Nicole Simonsen. 2019. "Gender-Based Violence Prevention Workshops: An Experimental Evaluation of Efficacy." *International Journal of Public Administration* 42(10):840-54.

Donnelly, Frank. 2019. *Exploring the U.S. Census.* Thousand Oaks, CA: Sage.

Doucet, Andrea and Natasha Mauthner. 2008. "Qualitative Interviewing and Feminist Research." Pp. 328-43 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Douglas, Jack D. 1985. *Creative Interviewing.* Beverly Hills, CA: Sage.

Drake, Robert E., Gregory J. McHugo, and Jeremy C. Biesanz. 1995. "The Test-Retest Reliability of Standardized Instruments Among Homeless Persons With Substance Use Disorders." *Journal of Studies on Alcohol* 56(2):161-67.

Drakulich, Kevin, Kevin H. Wozniak, John Hagan, and Devon Johnson. 2021. "Whose Lives Mattered? How White and Black Americans Felt About Black Lives Matter in 2016." *Law and Society Review* 55(2):227-51.

Drew, Paul. 2005. "Conversation Analysis." Pp. 71-102 in *Handbook of Language and Social Interaction,* edited by Kristine L. Fitch and Robert E. Sanders. Mahwah, NJ: Lawrence Erlbaum.

Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. 2021. "Affective Polarization, Local Contexts and Public Opinion in America." *Nature Human Behaviour* 5(1):28-38.

Drug Abuse Resistance Education (D.A.R.E.). 2021. D.A.R.E. America (https://dare.org/where-is-d-a-r-e/#USA).

Duckworth, Kenneth, John H. Halpern, Russell K. Schutt, and Christopher Gillespie. 2003. "Use of Schizophrenia as a Metaphor in U.S. News-papers." *Psychiatric Services* 54:1402–04.

Dum, Christopher P. 2016. *Exiled in America: Life on the Margins in a Residential Motel.* New York, NY: Columbia University Press.

Duncombe, Jean and Julie Jessop. 2002. "'Doing Rapport' and the Ethics of 'Faking Friendship.'" Pp. 107-22 in *Ethics in Qualitative Research,* edited by Melanie Mauthner, Maxine Birch, Julie Jessop, and Tina Miller. Thousand Oaks, CA: Sage.

Duneier, Mitchell. 1999. *Sidewalk.* New York, NY: Farrar, Strauss, and Giroux.

Duneier, Mitchell. 2016. *Ghetto: The Invention of a Place, the History of an Idea.* New York, NY: Farrar, Strauss, and Giroux.

Durkheim, Émile. 1893/2014. *The Division of Labor in Society.* Edited and with an Introduction by Steven Lukes. Translation by W. D. Halls. New York, NY: Free Press.

Durkheim, Émile. 1951. *Suicide.* New York, NY: Free Press.

Durkheim, Émile. 1966. *Suicide: A Study in Sociology.* Translated by John A. Spaulding and George Simpson. New York, NY: Free Press.

Dworznik-Hoak, Gretchen. 2020. "Making Sense of Harvey: An Exploration of How Journalists Find Meaning in Disaster." *Newspaper Research Journal* 41(2):160-78.

Dykema, Jennifer and Nora Cate Schaeffer. 2000. "Events, Instruments, and Reporting Errors." *American Sociological Review* 65:619-29.

Eckholm, Erik. 2006. "Report on Impact of Federal Benefits on Curbing Poverty Reignites a Debate." *The New York Times,* February 18, A8.

Edmonds, W. Alex and Thomas D. Kennedy. 2016. *An Applied Guide to Research Designs: Quantitative, Qualitative, and Mixed Methods,* 2nd ed. Thousand Oaks, CA: Sage.

Edwards, Frank, Hedwig Lee, and Michael Esposito. 2019. "Risk of Being Killed by Police Use of Force in the United States by Age, Race–Ethnicity, and Sex." *Proceedings of the National Academy of Sciences of the United States of America* 116(34):16793-98.

Egan, Patrick. 2012. "The Declining Culture of Guns and Violence in the United States." The Monkey Cage Blog, posted July 21 (http://themonkeycage.org/2012/07/21/the-declining-culture-of-guns-and-violence-in-the-united-states/).

Elder, Keith, Sudha Xirasagar, Nancy Miller, Shelly Ann Bowen, Saundra Glover, and Crystal Piper. 2007. "African Americans' Decisions Not to Evacuate New Orleans Before Hurricane Katrina: A Qualitative Study." *American Journal of Public Health* 97(Suppl. 1):S124-S129.

Elliott, Jane, Janet Holland, and Rachel Thomson. 2008. "Longitudinal and Panel Studies." Pp. 228-48 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Ellis, Carolyn. 1995. "Emotional and Ethical Quagmires in Returning to the Field." *Journal of Contemporary Ethnography* 24:68-98.

Ember, Carol R. and Melvin Ember. 2011. *A Basic Guide to Cross-Cultural Research.* New Haven, CT: Human Relations Area Files, Yale University (www.yale.edu/hraf/guides.htm).

Emerson, Robert M. (Ed.). 1983. *Contemporary Field Research.* Prospect Heights, IL: Waveland.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 1995. *Writing Ethnographic Fieldnotes.* Chicago, IL: University of Chicago Press.

Emmel, Nick and Andrew Clark. 2011. "Learning to Use Visual Methodologies in Our Research: A Dialogue Between Two Researchers." *Forum: Qualitative Social Research* 12(1): Article 36 (http://nbn-resolving.de/urn:nbn:de:0114-fqs1101360).

Emmison, Michael, Philip Smith, and Margery Mayall. 2012. *Researching the Visual,* 2nd ed. Thousand Oaks, CA: Sage.

Ennett, Susan T., Karl E. Bauman, Andrea Hussong, Robert Faris, Vangie A. Foshee, and Li Cai. 2006. "The Peer Context of Adolescent Substance Use: Findings From Social Network Analysis." *Journal of Research on Adolescence* 16:159-86.

Ennett, Susan T., Nancy S. Tobler, Christopher L. Ringwalt, and Robert L. Flewelling. 1994. "How Effective Is Drug Abuse Resistance Education? A Meta-Analysis of Project DARE Outcome Evaluations." *American Journal of Public Health* 84(9):1394-401.

Erikson, Kai T. 1966. *Wayward Puritans: A Study in the Sociology of Deviance.* New York, NY: Wiley.

Erikson, Kai T. 1967. "A Comment on Disguised Observation in Sociology." *Social Problems* 12:366-73.

Ettelt, Stefanie, Nicholas Mays, and Ellen Nolte. 2013. "Policy–Research Linkage: What We Have Learned From Providing a Rapid Response Facility for International Healthcare Comparisons to the Department of Health in England." *Evidence & Policy* 9:245-54.

Eurostat. 2003. *ECHP UDB Manual: European Community Household Panel Longitudinal Users' Database.* Brussels, Belgium: European Commission.

Exum, M. Lyn. 2002. "The Application and Robustness of the Rational Choice Perspective in the Study of Intoxicated and Angry Intentions to Aggress." *Criminology* 40:933-66.

Fallon, Kathleen M., Liam Swiss, and Jocelyn Viterna. 2012. "Resolving the Democracy Paradox: Democratization and Women's Legislative Representation in Developing Nations, 1975 to 2009." *American Sociological Review* 77(3):380-408.

Fawcett, Barbara and Rosalie Pockett. 2015. *Turning Ideas into Research: Theory, Design and Practice.* Thousand Oaks, CA: Sage.

Fears, Darryl. 2002. "For Latinos in U.S., Race Not Just Black or White." *The Boston Globe,* December 30, A3.

Federal Bureau of Investigation. 2021b. "Trend of Homicide from 1986 to 2019." *Crime Data Explorer,* September 18 (https://crime-data-explorer.app.cloud.gov/pages/explorer/crime/crime-trend).

Federal Bureau of Investigation. n.d. "NIBRS Overview" (https://ucr.fbi.gov/nibrs-overview).

Federal Register. 2017. Federal Policy for the Protection of Human Subjects, 82(12, Part IX), 49 CFR Part 11, January 19:7149–7274 (https://www.gpo.gov/fdsys/pkg/FR-2017-01-19/html/2017-01058.htm).

Fedyuk, Olena. 2012. "Images of Transnational Motherhood: The Role of Photographs in Measuring Time and Maintaining Connections Between Ukraine and Italy." *Journal of Ethnic and Migration Studies* 38:279-300.

Fenno, Richard F. Jr. 1978. *Home Style: House Members in Their Districts.* Boston: Little, Brown.

Fenton, Steve. 1996. "Counting Ethnicity: Social Groups and Official Categories." Pp. 143-65 in *Interpreting Official Statistics,* edited by Ruth Levitas and Will Guy. New York, NY: Routledge.

Ferdinand, Keith C. 2021. "Overcoming Barriers to COVID-19 Vaccination in African Americans: The Need for Cultural Humility." *American Journal of Public Health* 111(4):586-88.

Ferguson, Karen J. 1998. "Caught in 'No Man's Land': The Negro Cooperative Demonstration Service and the Ideology of Booker T. Washington, 1900–1918." *Agricultural History* 72(1):33-54.

Ferguson, Kristin M., Kimberly Bender, Sanna J. Thompson, Elaine M. Maccio, and David Pollio. 2012. "Employment Status and Income Generation Among Homeless Young Adults: Results From a Five-City, Mixed-Methods Study." *Youth & Society* 44:385-407.

Fetterman, David M. 2019. *Ethnography.* 4th ed. New York, NY: Sage.

File, Thom. 2013. "Computer and Internet Use in the United States." Pp. 20-569 in *Current Population Survey Reports.* Washington, DC: U.S. Census Bureau.

Filindra, Alexandra and Noah Kaplan. 2017. "Testing Theories of Gun Policy Preferences Among Blacks, Latinos, and Whites in America*." *Social Science Quarterly* 98(2):413-28.

Fink, Arlene. 2005. *Conducting Research Literature Reviews: From the Internet to Paper,* 2nd ed. Thousand Oaks, CA: Sage.

Finkel, Eli J., Paul W. Eastwick, and Harry T. Reis. 2017. "Replicability and Other Features of a High-Quality Science: Toward a Balanced and Empirical Approach." *Journal of Personality and Social Psychology* 113:244-53.

Firebaugh, Glenn. 2008. *Seven Rules for Social Research.* Princeton, NJ: Princeton University Press.

Fischer, Claude. 2009. "The 2004 GSS Finding of Shrunken Social Networks: An Artifact?" *American Sociological Review* 74(4):657-69.

Fischer, Constance T. and Frederick J. Wertz. 2002. "Empirical Phenomenological Analyses of Being Criminally Victimized." Pp. 275-304 in *The Qualitative Researcher's Companion,* edited by A. Michael Huberman and Matthew B. Miles. Thousand Oaks, CA: Sage.

Fisher, Celia B. and Andrea E. Anushko. 2008. "Research Ethics in Social Science." Pp. 94-109 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Fligstein, Neil, Orestes P. Hastings, and Adam Goldstein. 2017. "Keeping Up With the Joneses: How Households Fared in the Era of High Income Inequality and the Housing Price Bubble, 1999–2007. *Socius: Sociological Research for a Dynamic World* 3:1-15.

Forero, Juan. 2000a. "Census Takers Say Supervisors Fostered Filing of False Data." *The New York Times,* July 28, A21.

Forero, Juan. 2000b. "Census Takers Top '90 Efforts in New York City, With More to Go." *The New York Times,* June 12, A29.

Fowler, Floyd J. 1988. *Survey Research Methods,* rev. ed. Newbury Park, CA: Sage.

Fowler, Floyd J. 1995. *Improving Survey Questions: Design and Evaluation.* Thousand Oaks, CA: Sage.

Fox, Chris, Robert Grimm, and Rute Caldeira. 2016. *An Introduction to Evaluation.* Thousand Oaks, CA: Sage.

Fox, Nick and Chris Roberts. 1999. "GPs in Cyberspace: The Sociology of a 'Virtual Community.'" *The Sociological Review* 47:643-69.

Fox-Williams, Brittany N. 2019. "The Rules of (Dis)Engagement: Black Youth and Their Strategies for Navigating Police Contact." *Sociological Forum* 34(1):115-37.

Frank, David John, Ann Hironaka, and Evan Schofer. 2000. "The Nation-State and the Natural Environment Over the Twentieth Century." *American Sociological Review* 65:96-116.

Frankel, Laura Lazarus and D. Sunshine Hillygus. 2013. "Looking Beyond Demographics: Panel Attrition in the ANES and GSS." *Political Analysis* 22(3):336-53.

Frankfort-Nachmias, Chava and Anna Leon-Guerrero. 2006. *Social Statistics for a Diverse Society.* 4th ed. Thousand Oaks, CA: Sage.

Frankfort-Nachmias, Chava, Anna Leon-Guerrero, and Georgiann Davis. 2020. *Social Statistics for a Diverse Society.* 9th ed. Thousand Oaks, CA: Sage.

Franklin, Mark N. 1996. "Electoral Participation." Pp. 216-35 in *Comparing Democracies: Elections and Voting in Global Perspective,* edited by Lawrence LeDuc, Richard G. Niemi, and Pippa Norris. Thousand Oaks, CA: Sage.

Frazier, Cleothia. 2020. "It's More Than Just News: Print Media, the Tuskegee Syphilis Study and Collective Memory Among African Americans." *Journal of Historical Sociology* 33(3):280-96.

Freshman, Audrey. 2012. "Financial Disaster as a Risk Factor for Posttraumatic Stress Disorder: Internet Survey of Trauma in Victims of the Madoff Ponzi Scheme." *Health & Social Work* 37:39-48.

Frohmann, Lisa. 2005. "The Framing Safety Project: Photographs and Narratives by Battered Women." *Violence Against Women* 11:1396-419.

Funk, Cary, Brian Kennedy, and Courtney Johnson. 2020. *Trust in Medical Science Has Grown in the U.S., but Mainly Among Democrats.* Washington, DC: Pew Research Center.

Gaiser, Ted J. and Anthony E. Schreiner. 2009. *A Guide to Conducting Online Research.* Thousand Oaks, CA: Sage.

Gall, Carlotta. 2003. "Armed With Pencils, Army of Census Workers Fans Out Into Afghan Outback." *The New York Times,* July 15, A4.

Gallup. 2011. *Election Polls—Accuracy Record in Presidential Elections* (http://www.gallup.com/poll/9442/Election-Polls-Accuracy-Record-Presidential-Elections.aspx?version=print).

Gans, Curtis. 2008. "2008 Primary Turnout Falls Just Short of Record Nationally, Breaks Records in Most States." *AU News,* May 19. Washington, DC: Center for the Study of the American Electorate, American University (www.american.edu/media).

GAO@100 n.d. "Decennial Census." (https://www.gao.gov/highrisk/decennial-census).

García-Rapp, Florencia. 2019. "Trivial and Normative? Online Fieldwork Within Youtube's Beauty Community." *Journal of Contemporary Ethnography* 48(5):619-44.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology.* Englewood Cliffs, NJ: Prentice Hall.

Geertz, Clifford. 1973. "Thick Description: Toward an Interpretive Theory of Culture." Pp. 3-30 in *The Interpretation of Cultures,* edited by Clifford Geertz. New York, NY: Basic Books.

Gehl, Robert W. 2016. "Power/Freedom on the Dark Web: A Digital Ethnography of the Dark Web Social Network." *New Media and Society* 18(7):1219-35.

Gerth, Matthias A. and Gabriele Siegert. 2012. "Patterns of Consistence and Constriction: How News Media Frame the Coverage of Direct Democratic Campaigns." *American Behavioral Scientist* 56:279-99.

Gil, Paul. 2017. "The Best Search Engines of 2017," *Lifewire,* September 11 (https://www.lifewire.com/best-search-engines-2483352).

Gilchrist, Valerie J. and Robert L. Williams. 1999. "Key Informant Interviews." Pp. 71-88 in *Doing Qualitative Research,* 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Giles, Jim. 2005. "Internet Encyclopaedias Go Head to Head." *Nature* 438:900-01.

Gill, Hannah E. 2004. "Finding a Middle Ground Between Extremes: Notes on Researching Transnational Crime and Violence." *Anthropology Matters Journal* 6:1-9.

Gill, Richard T., Nathan Glazer, and Stephan A. Thernstrom. 1992. *Our Changing Population.* Englewood Cliffs, NJ: Prentice Hall.

Gilligan, Carol. 1988. "Adolescent Development Reconsidered." Pp. vii-xxxix in *Mapping the Moral Domain,* edited by Carol Gilligan, Janie Victoria Ward, and Jill McLean Taylor. Cambridge, MA: Harvard University Press.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. "Detecting Influenza Epidemics Using Search Engine Query Data." *Nature* 457(February 19):1012-15.

Glaser, Barney G. and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research.* London: Weidenfeld and Nicholson.

Glover, Judith. 1996. "Epistemological and Methodological Considerations in Secondary Analysis." Pp. 28-38 in *Cross-National Research Methods in the Social Sciences,* edited by Linda Hantrais and Steen Mangen. New York, NY: Pinter.

Glueck, Sheldon and Eleanor Glueck. 1950. *Unraveling Juvenile Delinquency.* New York, NY: Commonwealth Fund.

Gobo, Giampietro. 2008. "Re-Conceptualizing Generalization: Old Issues in a New Frame." Pp. 193-213 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Goertz, Gary. 2006. *Social Science Concepts: A User's Guide.* Princeton, NJ: Princeton University Press.

Goffman, Erving. 1961. *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates.* Garden City, NY: Doubleday.

Goffman, Erving. 1963. *Stigma: Notes on the Management of Spoiled Identity.* Englewood, NJ: Prentice Hall/Spectrum.

Goffman, Erving. 1979. *Gender Advertisements.* London: Macmillan.

Gold, Raymond. 1958. "Roles in Sociological Field Observation." *Social Forces* 36:217-23.

Golder, Scott A. and Michael W. Macy. 2011. "Diurnal and Seasonal Mood Vary With Work, Sleep, and Daylength Across Diverse Cultures." *Science* 333:1878.

Golder, Scott A. and Michael W. Macy. 2014. "Digital Footprints: Opportunities and Challenges for Online Social Research." *Annual Review of Sociology* 40:129-52.

Goldfinger, Stephen M. and Russell K. Schutt. 1996. "Comparison of Clinicians' Housing Recommendations and Preferences of Homeless Mentally Ill Persons." *Psychiatric Services,* 47:413-15.

Goldfinger, Stephen M., Russell K. Schutt, Larry J. Seidman, Winston M. Turner, Walter E. Penk, and George S. Tolomiczenko. 1996. "Self-Report and Observer Measures of Substance Abuse Among Homeless Mentally Ill Persons in the Cross-Section and Over Time." *The Journal of Nervous and Mental Disease* 184(11):667-72.

Goleman, Daniel. 1993a. "Placebo Effect Is Shown to Be Twice as Powerful as Expected." *The New York Times,* August 17, C3.

Goleman, Daniel. 1993b. "Pollsters Enlist Psychologists in Quest for Unbiased Results." *The New York Times,* September 7, C1, C11.

Goode, Erich. 2002. "Sexual Involvement and Social Research in a Fat Civil Rights Organization." *Qualitative Sociology* 25:501-34.

Goodnough, Abby. 2010. "A Wave of Addiction and Crime, With the Medicine Cabinet to Blame." *The New York Times,* September 22.

Gordon, Raymond. 1992. *Basic Interviewing Skills.* Itasca, IL: Peacock.

Gough, David, Sandy Oliver, and James Thomas. 2017. *An Introduction to Systematic Reviews,* 2nd ed. Thousand Oaks, CA: Sage.

Graduate Program in Applied Sociology. 1990. *Handbook for Thesis Writers.* Boston, MA: University of Massachusetts.

Grady, John. 1996. "The Scope of Visual Sociology." *Visual Sociology* 11:10-24.

Grant, Bridget F., Deborah A. Dawson, Frederick S. Stinson, S. Patricia Chou, Mary C. Dufour, and Roger P. Pickering. 2004. "The 12-Month Prevalence and Trends in DSM-IV Alcohol Abuse and Dependence: United States, 1991–1992 and 2001–2002." *Drug and Alcohol Dependence* 74:223-34.

Greenberg, David, Mark Shroder, and Mathew Onstott. 1999. *The Digest of Social Experiments,* 2nd ed. Washington, DC: Urban Institute Press.

Greenhalgh, Trisha, Manuel B. Schmid, Thomas Czypionka, Dirk Bassler, and Laurence Gruer. 2020. "Face Masks for the Public During the Covid-19 Crisis." *The BMJ* 369(April):1-4.

Grey, Robert J. Jr. 2005. "Jury Service: It's a Privilege." American Bar Association (www.abanet.org/media/releases/opedjuror 2.html).

Grieco, Elizabeth M., Yesenia D. Acosta, G. Patricia de la Cruz, Christine Gambino, Thomas Gryn, Luke J. Larsen, Edward N. Trevelyan, and Nathan P. Walters. 2012. *The Foreign-Born Population in the United States: 2010.* Washington, DC: U.S. Census Bureau.

Griffin, Larry J. 1992. "Temporality, Events, and Explanation in Historical Sociology: An Introduction." *Sociological Methods & Research* 20:403-27.

Griffin, Larry J. 1993. "Narrative, Event-Structure Analysis, and Causal Interpretation in Historical Sociology." *American Journal of Sociology* 98(March):1094-133.

Grinnell, Frederick. 1992. *The Scientific Attitude,* 2nd ed. New York, NY: Guilford.

Grissom, Brandi. 2011. "Proposals Could Make It Harder to Leave Prison." *The New York Times,* March 12.

Groeniger, Joost Oude, Kjell Noordzij, Jeroen van der Waal, and Willem de Koster. 2021. "Dutch COVID-19 Lockdown Measures Increased Trust in Government and in Science: A Difference-in-Differences Analysis." *Social Science & Medicine* 275:113819.

Groves, Robert M. 1989. *Survey Errors and Survey Costs.* New York, NY: Wiley.

Groves, Robert M. and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys.* New York, NY: Wiley.

Groves, Robert M. and Robert L. Kahn. 1979. *Surveys by Telephone: A National Comparison With Personal Interviews.* New York, NY: Academic Press. As adapted in Delbert C. Miller, 1991. *Handbook of Research Design and Social Measurement,* 5th ed. Newbury Park, CA: Sage.

Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. "Leverage-Salience Theory of Survey Participation: Description and an Illustration." *Public Opinion Quarterly* 64:299-308.

Gruber, Maria, Jakob-Moritz Eberl, Fabienne Lind, and Hajo G. Boomgaarden. 2021. "Qualitative Interviews With Irregular Migrants in Times of Covid-19: Recourse to Remote Interview Techniques as a Possible Methodological Adjustment." *Forum Qualitative Sozialforschung* 22(1):1-15.

Gruenewald, Paul J., Andrew J. Treno, Gail Taff, and Michael Klitzner. 1997. *Measuring Community Indicators: A Systems Approach to Drug and Alcohol Problems.* Thousand Oaks, CA: Sage.

Guba, Egon G. and Yvonna S. Lincoln. 1989. *Fourth Generation Evaluation.* Newbury Park, CA: Sage.

Guba, Egon G. and Yvonna S. Lincoln. 1994. "Competing Paradigms in Qualitative Research." Pp. 105-17 in *Handbook of Qualitative Research,* edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Guba, Egon, Lincoln Yvonna, and Lynham Sue. 2017. "Paradigmatic Controversies, Contradictions, and Emerging Confluences" edited by Denzin and Lincoln. 5th ed. The SAGE Handbook of Qualitative Research.

Gubrium, Jaber F. and James A. Holstein. 1997. *The New Language of Qualitative Method.* New York: Oxford University Press.

Guterbock, Thomas M. 2008. *Strategies and Standards for Reaching Respondents in an Age of New Technology.* Presentation to the Harvard Program on Survey Research Spring Conference, New Technologies and Survey Research. Cambridge, MA: Institute of Quantitative Social Science, Harvard University, May 9.

Hacker, Karen, Jessica Collins, Leni Gross-Young, Stephanie Almeida, and Noreen Burke. 2008. "Coping With Youth Suicide and Overdose: One Community's Efforts to Investigate, Intervene, and Prevent Suicide Contagion." *Crisis* 29:86-95.

Hacker, Karen. 2013. *Community-Based Participatory Research.* Thousand Oaks, CA: Sage.

Hadaway, C. Kirk, Penny Long Marler, and Mark Chaves. 1993. "What the Polls Don't Show: A Closer Look at U.S. Church Attendance." *American Sociological Review* 58(6):741-52.

Hafner, Katie. 2004. "For Some, the Blogging Never Stops." *The New York Times,* May 27, E1, E7.

Hage, Jerald and Barbara Foley Meeker. 1988. *Social Causality.* Boston, MA: Unwin Hyman.

Hakim, Catherine. 1982. *Secondary Analysis in Social Research: A Guide to Data Sources and Methods With Examples.* London: George Allen & Unwin.

Hakimzadeh, Shirin and D'Vera Cohn. 2007. *English Usage Among Hispanics in the United States.* Washington, DC: Pew Hispanic Center (http://pewhispanic.org/files/reports/82.pdf).

Hallinan, Maureen T. 1997. "The Sociological Study of Social Change." *American Sociological Review* 62:1-11.

Hammersley, Martyn. 2008. "Assessing Validity in Social Research." Pp. 42-53 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Hammersley, Martyn and Anna Traianou. 2012. *Ethics in Qualitative Research: Controversies and Contexts.* Thousand Oaks, CA: Sage.

Hampton, Keith N. 2003. "Grieving for a Lost Network: Collective Action in a Wired Suburb." *The Information Society* 19:417-28.

Hampton, Keith N. and Barry Wellman. 1999. "Netville On-line and Off-Line: Observing and Surveying a Wired Suburb." *American Behavioral Scientist* 43:475-92.

Hampton, Keith N. and Barry Wellman. 2000. "Examining Community in the Digital Neighborhood: Early Results From Canada's Wired Suburb." Pp. 475-92 in *Digital Cities: Technologies, Experiences, and Future Perspectives,* edited by Toru Ishida and Katherine Isbister. Berlin, Germany: Springer-Verlag.

Hampton, Keith N. and Barry Wellman. 2001. "Long Distance Community in the Network Society." *American Behavioral Scientist* 45:476-95.

Hampton, Keith N. and Neeti Gupta. 2008. "Community and Social Interaction in the Wireless City: Wi-Fi Use in Public and Semi-Public Spaces." *New Media & Society* 10(6):831-50.

Hampton, Keith N., Lauren Sessions Goulet, Eun Ja Her, and Lee Rainie. 2009. *Social Isolation and New Technology: How the Internet and Mobile Phones Impact Americans' Social Networks.* Washington, DC: Pew Internet & American Life Project (http://www.pewinternet.org/2009/11/04/social-isolation-and-new-technology/).

Hampton, Keith N., Lauren Sessions Goulet, Lee Rainie, and Kristen Purcell. 2011. *Social Networking Sites and Our Lives: How People's Trust, Personal Relationships, and Civic and Political Involvement Are Connected to Their Use of Social Networking Sites and Other Technologies.* Washington, DC: Pew Internet & American Life Project (http://pewinternet.org/Reports/2011/Technologyand-social-networks.aspx).

Haney, C., C. Banks, and Philip G. Zimbardo. 1973. "Interpersonal Dynamics in a Simulated Prison." *International Journal of Criminology and Penology* 1:69-97.

Hantrais, Linda and Steen Mangen. 1996. "Method and Management of Cross-National Social Research." Pp. 1-12 in *Cross-National Research Methods in the Social Sciences,* edited by Linda Hantrais and Steen Mangen. New York, NY: Pinter.

Hard, Stephen F., James M. Conway, and Antonia C. Moran. 2006. "Faculty and College Student Beliefs About the Frequency of Student Academic Misconduct." *The Journal of Higher Education* 77:1058-80.

Harding, David J. 2007. "Cultural Context, Sexual Behavior, and Romantic Relationships in Disadvantaged Neighborhoods." *American Sociological Review* 72:341-64.

Hargraves, Lee, Russell K. Schutt, Daniel T. O'Bien, Dan Dooley, Floyd (Jack) Fowler, Ryan Qi Wang, Alina Ristea, Anthony Roman, Mehrnaz Amiri, Sage Gibbons, Hannah Grabowski, and Nikola Kovacevic. 2021. *Living in Boston During COVID-19: Vaccination Planning and Hesitancy. Report #5.*

Harley, Anne and Jonathan Langdon. 2018. "Ethics and Power in Visual Research." Pp. 188-202 in *The SAGE Handbook of Qualitative Research Ethics,* edited by R. Iphofen & M. Tolich. London: Sage.

Harris, David R. and Jeremiah Joseph Sim. 2002. "Who Is Multiracial? Assessing the Complexity of Lived Race." *American Sociological Review* 67:614-27.

Harris, Jeffrey E. 2021. "COVID-19 Incidence and Hospitalization Rates Were Inversely Related to Vaccination Coverage Among the 112 Most Populous Counties in the United States." *Health Policy and Technology* (https://pubmed.ncbi.nlm.nih.gov/34868833/).

Harris, Poll. 2017. "Rejecting Labels and Sparking an Identity Revolution, Millennials Refuse to be Trumped" (http://www.theharrispoll.com/politics/Rejecting-Labels-Identity-Revolution.html).

Hart, Chris. 1998. *Doing a Literature Review: Releasing the Social Science Research Imagination.* London: Sage.

Hasin, Deborah S., Frederick S. Stinson, Elizabeth Ogburn, and Bridget F. Grant. 2007. "Prevalence, Correlates, Disability, and Comorbidity of *DSM-IV* Alcohol Abuse and Dependence in the United States: Results From the National Epidemiologic Survey on Alcohol and Related Conditions." *Archives of General Psychiatry* 64:830-42.

Hawkins, Donald N., Paul R. Amato, and Valarie King. 2007. "Nonresident Father Involvement and Adolescent Well-Being: Father Effects or Child Effects?" *American Sociological Review* 72: 990-1010.

Healy, Jack. 2017. "For One Family in Houston, an Overwhelming Start to Recovery." *The New York Times,* September 2 (https://www.nytimes.com/interactive/2017/09/01/upshot/cost-of-hurricaneharvey-only-one-storm-comes-close.html).

Hearld, Kristine R., Danielle Wu, and Henna Budhwani. 2021. "HIV Testing Among Muslim Women in the United States: Results of a National Sample Study." *Health Equity* 5(1):17-22.

Heath, Christian and Paul Luff. 2008. "Video and the Analysis of Work and Interaction." Pp. 493–505 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Heaton, Janet. 2008. "Secondary Analysis of Qualitative Data." Pp. 506-35 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Heckathorn, Douglas D. 1997. "Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations." *Social Problems* 44:174-99.

Heckathorn, Douglas D. 2002. "Respondent-Driven Sampling II: Deriving Valid Population Estimates From Chain-Referral Samples of Hidden Populations." *Social Problems* 49:11-34.

Heckman, James, Neil Hohmann, and Jeffrey Smith. 2000. "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment." *The Quarterly Journal of Economics* 115:651-94.

Hedström, Peter and Richard Swedberg (Eds.). 1998. *Social Mechanisms: An Analytical Approach to Social Theory.* Cambridge: Cambridge University Press.

Heise, David R. 2010. *Surveying Cultures: Discovering Shared Conceptions and Sentiments.* Hoboken, NJ: Wiley.

Heitman, Stephanie. 2021. "What Happens in an Internet Minute in 2021: 88 Fascinating Online Stats." *LOCALiQ,* May 21.

Hemmer, Lynn and D. Scott Elliff. 2020. "Leaders in Action: The Experiences of Seven Texas Superintendents Before, During, and After Hurricane Harvey." *Educational Management Administration and Leadership* 48(6):964-85.

Herek, Gregory. 1995. "Developing a Theoretical Framework and Rationale for a Research Proposal." Pp. 85-91 in *How to Write a Successful Research Grant Application: A Guide for Social and Behavioral Scientists,* edited by Willo Pequegnat and Ellen Stover. New York, NY: Plenum Press.

Heritage, John. n.d. *Conversation Analysis. Chapter in E-source, Social and Behavioral Sciences Research.* Washington, DC: Office of Behavioral and Social Sciences Research, National Institutes of Health (www.esourceresearch.org/eSourceBook/ConversationAnalysis/10TranscriptionSymbols/tabid/531/Default.aspx).

Hesse-Biber, Sharlene. 1989. "Eating Problems and Disorders in a College Population: Are College Women's Eating Problems a New Phenomenon?" *Sex Roles* 20:71-89.

Hewson, Claire, Carl Vogel, and Dianna Laurent. 2016. *Internet Research Methods,* 2nd ed. Thousand Oaks, CA: Sage.

Hicks, Lorna. 2013. *The Regulations—SBE.* Miami, FL: Collaborative Institutional Training Initiative at the University of Miami (https://www.citiprogram.org/).

Hill, Joan and Jennifer Reichert. 2021. 2020 Census Program for Evaluations, Experiments, and Assessments. Washington, D.C.: U.S. Bureau of the Census (https://www.census.gov/newsroom/blogs/random-samplings/2021/09/2020-census-program-for-evaluations-experiments-and-assessments.html).

The Hippocratic Oath. n.d. Charlottesville: University of Virginia, Historica Medicina (http://exhibits.hsl.virginia.edu/antiqua/hippocrates/).

Hirsch, Kathleen. 1989. *Songs From the Alley.* New York, NY: Doubleday.

Hirschel, David, Eve Buzawa, April Pattavina, and Don Faggiani. 2008. "Domestic Violence and Mandatory Arrest Laws: To What Extent Do They Influence Police Arrest Decisions?" *Journal of Criminal Law & Criminology* 98(1):255-98.

Hjerm, Mikael, Maureen A. Eger, Andrea Bohman, and Filip Fors Connolly. 2020. "A New Approach to the Study of Tolerance: Conceptualizing and Measuring Acceptance, Respect, and Appreciation of Difference." *Social Indicators Research* 147(3):897-919.

Ho, D. Y. F. 1996. "Filial Piety and Its Psychological Consequences." Pp. 155-65 in *Handbook of Chinese Psychology,* edited by M. H. Bond. Hong Kong: Oxford University Press.

Hoberman, John. 2012. "Examining Tuskegee." *Society* 49(3):292-95.

Hofling, Charles K., Eveline M. Brotzman, Sarah M. Dalrymple, Nancy M. Graves, and Chester M. Pierce. 1966. "An Experimental Study of Nurse-Physician Relations." *Journal of Nervous and Mental Disease* 143:171-80.

Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone Versus Face-to-Face Interviewing of National Probability Samples With Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias." *Public Opinion Quarterly* 60:58-88.

Hollingsworth, T. H. 1972. "The Importance of the Quality of Data in Historical Demography." Pp. 71-86 in *Population and Social Change,* edited by D. V. Glass and Roger Revelle. London: Edward Arnold.

Holmes, Steven A. 2000. "Stronger Response by Minorities Helps Improve Census Reply Rate." *The New York Times,* May 4, A1, A22.

Holmes, Steven A. 2001a. "Census Officials Ponder Adjustments Crucial to Redistricting." *The New York Times,* February 12, A17.

Holmes, Steven A. 2001b. "The Confusion Over Who We Are." *The New York Times,* June 3, WK 1.

Horney, Julie, D. Wayne Osgood, and Ineke Haen Marshall. 1995. "Criminal Careers in the Short-Term: Intra-Individual Variability in Crime and Its Relation to Local Life Circumstances." *American Sociological Review* 60:655-73.

Horowitz, Carol R., Mimsie Robinson, and Sarena Seifer. 2009. "Community-Based Participatory Research From the Margin to the Mainstream: Are Researchers Prepared?" *Circulation* 119:2633-42.

Houtzager, Peter P. and Arnab K. Acharya. 2011. "Associations, Active Citizenship, and the Quality of Democracy in Brazil and Mexico." *Theory and Society* 40:1-36.

Howell, James C. 2003. *Preventing and Reducing Juvenile Delinquency: A Comprehensive Framework.* Thousand Oaks, CA: Sage.

Hoyle, Carolyn and Andrew Sanders. 2000. "Police Response to Domestic Violence: From Victim Choice to Victim Empowerment." *British Journal of Criminology* 40:14-26.

Hrobjartsson, Asbjorn and Peter C. Gotzsche. 2001. "Is the Placebo Powerless? An Analysis of Clinical Trials Comparing Placebo With No Treatment." *New England Journal of Medicine* 344:1594-1602.

Hswen, Yulin, Xiang Xu, Anna Hing, Jared B. Hawkins, John S. Brownstein, and Gilbert C. Ghee. 2021. "Association of "#covid19" Versus "#chinesevirus" With Anti-Asian Sentiments on Twitter: March 9-23, 2020. *American Journal of Public Health* 111(5):956-864 (https://www.nytimes.com/2020/04/06/nyregion/coronavirus-nyc-child-abuse.html).

Hu, Winnie. 2014. "Severe Cold Moves New York's Homeless to Seek Help." *The New York Times,* January 28, A19.

Huberman, A. Michael and Matthew B. Miles. 1994. "Data Management and Analysis Methods." Pp. 428–444 in *Handbook of Qualitative Research,* edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Huff, Darrell. 1954. *How to Lie With Statistics.* New York, NY: Norton.

Huff, Jessica, Charles M. Katz, and E. C. Hedberg. 2020. "A Randomized Controlled Trial of the Impact of Body-Worn Camera Activation on the Outcomes of Individual Incidents." *Journal of Experimental Criminology.*

Humes, Karen R., Nicholas A. Jones, and Roberto R. Ramirez. 2011. *Overview of Race and Hispanic Origin: 2010.* C2010BR-02. Washington, DC: U.S. Census Bureau.

Humphrey, Nicholas. 1992. *A History of the Mind: Evolution and the Birth of Consciousness.* New York, NY: Simon & Schuster.

Humphreys, Laud. 1970. *Tearoom Trade: Impersonal Sex in Public Places.* Chicago, IL: Aldine de Gruyter.

Humphries, Courtney. 2011. "Deeply Conflicted: How Can We Insulate Ourselves From Conflicts of Interest? The Most Popular Solution—Disclosing Them—Turns Out Not to Help." *Boston Globe,* May 15, K1, K3.

Hunt, Morton. 1985. *Profiles of Social Research: The Scientific Study of Human Interactions.* New York, NY: Russell Sage Foundation.

Huston, Patricia and C. David Naylor. 1996. "Health Services Research: Reporting on Studies Using Secondary Data Sources." *Canadian Medical Association Journal* 155:1697-702.

Hyvärinen, Matti. 2008. "Analyzing Narratives and Story-Telling." Pp. 447-60 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Iezzoni, Lisa I. 1997. "How Much Are We Willing to Pay for Information About Quality of Care?" *Annals of Internal Medicine* 126:391-93.

Innstrand, Siw Tone, Geir Arild Espries, and Reidar Mykletun. 2004. "Job Stress, Burnout and Job Satisfaction: An Intervention Study of Staff Working With People With Intellectual Disabilities." *Journal of Applied Research in Intellectual Disabilities* 17:119-26.

Internet Live Stats. n.d. "Total Number of Websites." Retrieved August 31, 2021 (https://www.internetlivestats.com/total-number-of-websites/).

Internet World Stats. 2022. "Internet World Stats: Usage and Population Statistics." Retrieved April 30, 2022 (http://www.internetworldstats.com/stats.htm).

Iphofen, R., & Tolich, M., eds. 2018. "Foundational Issues in Qualitative Research Ethics." Pp. 1–18 in *The SAGE Handbook of Qualitative Research Ethics.* Thousand Oaks, CA: Sage.

Irvine, Leslie. 1998. "Organizational Ethics and Fieldwork Realities: Negotiating Ethical Boundaries in Codependents Anonymous." Pp. 167-83 in *Doing Ethnographic Research: Fieldwork Settings.* Thousand Oaks, CA: Sage.

Israel, Mark. 2014. *Research Ethics and Integrity for Social Scientists: Beyond Regulatory Compliance,* 2nd ed. Thousand Oaks, CA: Sage.

Jackson, Ashley M. and Faruque Ahmed. 2020. "Assessing Characteristics of Unplanned School Closures That Occurred in the United States in Response to Hurricane Harvey in 2017." *Disaster Medicine and Public Health Preparedness* 14(1):125-29.

Jackson, C. 2021. *Topline and Methodology.* Washington, DC: Ipsos. SEAN-COVID-19-Survey-Summary_10-29-21.pdf (langerresearch.com) SEAN COVID-19 Survey Archive (parc.us.com)

Jacobs, Andrew. 2021. "The Struggle Against Vaccine Hesitancy Comes to Mississippi." *The New York Times,* April 9.

Jalbert, Sarah Kuck, William Rhodes, Christopher Flygare, and Michael Kane. 2010. "Testing Probation Outcomes in an Evidence-Based Practice Setting: Reduced Caseload Size and Intensive Supervision Effectiveness." *Journal of Offender Rehabilitation,* 49:233-53

James, Nalita and Hugh Busher. 2009. *Online Interviewing.* Thousand Oaks, CA: Sage.

Jang, Bohyun Joy, Megan E. Patrick, and Megan S. Schuler. 2017. "Substance Use Behaviors and the Timing of Family Formation During Young Adulthood." *Journal of Family Issues,* Online first.

Jarvis, Helen. 1997. "Housing, Labour Markets and Household Structure: Questioning the Role of Secondary Data Analysis in Sustaining the Polarization Debate." *Regional Studies* 31:521-31.

Jervis, Robert. 1996. "Counterfactuals, Causation, and Complexity." Pp. 309-16 in *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives,* edited by Philip E. Tetlock and Aaron Belkin. Princeton, NJ: Princeton University Press.

Jesnadum, Anick. 2000. "Researchers Fear Privacy Breaches with Online Research." *Digital Mass* (www.digitalmass.com/news/daily/09/15/researchers.html).

Johnson, David and Merry Bullock. 2009. "The Ethics of Data Archiving: Issues From Four Perspectives." Pp. 214-28 in *The Handbook of Social Research Ethics,* edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Jones, James H. 1993. *Bad Blood: The Tuskegee Syphilis Experiment.* New York, NY: Simon & Schuster.

Jones, Stephen R. G. 1992. "Was There a Hawthorne Effect?" *American Journal of Sociology* 98:451-68.

Kagay, Michael R. with Janet Elder. 1992. "Numbers Are No Problem for Pollsters. Words Are." *The New York Times,* October 9, E5.

Kale-Lostuvali, Elif. 2007. "Negotiating State Provision: State-Citizen Encounters in the Aftermath of the İzmit Earthquake." *The Sociological Quarterly* 48:745-67.

Kalton, Graham. 2020. *Introduction to Survey Sampling.* 2nd ed. Thousand Oaks, CA: Sage.

Kaplan, Fred. 2002. "NY Continues to See Plunge in Number of Felonies." *The Boston Globe,* April 15, A3.

Kato, Yuki, Catarina Passidomo, and Daina Harvey. 2014. "Political Gardening in Post-Disaster City: Lessons From New Orleans." *Urban Studies* 51:1833-49.

Kaufman, Sharon R. 1986. *The Ageless Self: Sources of Meaning in Late Life.* Madison, WI: University of Wisconsin Press.

Kaysen, Ronda. 2021. "Trying to Hold Back the Floodwaters." *The New York Times,* September 26, RE3.

Kee, Chad E. 2021. "The Impact of COVID-19: Graduate Students' Emotional and Psychological Experiences." *Journal of Human Behavior in the Social Environment* 31(1–4):476-88.

Keeter, Scott, Nick Hatley, Courtney Kennedy, and Arnold Lau. 2017. "What Low Response Rates Mean for Telephone Surveys." Pew Research Center, May 15 (http://www.pewresearch.org/2017/05/15/what-lowresponse-rates-mean-for-telephone-surveys/).

Keeter, Scott. 2008. *Survey Research and Cell Phones: Is There a Problem?* Presentation to the Harvard Program on Survey Research Spring Conference, New Technologies and Survey Research. Cambridge, MA: Institute of Quantitative Social Science, Harvard University, May 9.

Kemmis, Stephen and Robin McTaggart. 2005. "Participatory Action Research: Communicative Action and the Public Sphere." Pp. 559-603 in *The SAGE Handbook of Qualitative Research,* 3rd ed., edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Kemp, Simon. 2017. "The Global State of the Internet in 20107" (https://thenextweb.com/contributors/2017/04/11/current-global-state-internet/).

Kennedy, Brian, Alec Tyson, and Cary Funk. 2022. "Americans' Trust in Scientists, Other Groups Declines." Pew Research Center, Washington, DC, February 2022. Retrieved from (https://www.pewresearch.org/science/2022/02/15/americans-trust-in-scientists-other-groups-declines/).

Kennedy, Courtney and Hannah Hartig. 2019. "Response Rates in Telephone Surveys Have Resumed Their Decline." Pew Research Center. Retrieved November 18, 2021 (https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/).

Kennedy, David M., Anthony M. Piehl, and Anne A. Braga. 1996. "Youth Violence in Boston: Gun Markets, Serious Youth Offenders, and a Use-Reduction Strategy." *Law and Contemporary Problems* 59:147-96.

Kenney, Charles. 1987. "They've Got Your Number." *The Boston Globe Magazine,* August 30, 12, 46-56, 60.

Kershaw, David and Jerilyn Fair. 1976. *The New Jersey Income-Maintenance Experiment,* vol. 1. New York, NY: Academic Press.

Kershaw, Sarah. 2000. "In a Black Community, Mistrust of Government Hinders Census." *The New York Times,* May 16, A20.

Khoury, Rana B. 2020. "Hard-to-Survey Populations and Respondent-Driven Sampling: Expanding the Political Science Toolbox." *Perspectives on Politics* 18(2):509-26.

Kifner, John. 1994. "Pollster Finds Error on Holocaust Doubts." *The New York Times,* May 20, A12.

Kilger, Magnus. 2017. "Talent Stories in Youth Sports: Discursively Shared Narratives of Success." *Narrative Inquiry* 27:47-65.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Scientific Inference in Qualitative Research.* Princeton, NJ: Princeton University Press.

King, Miriam L. and Diana L. Magnuson. 1995. "Perspectives on Historical U.S. Census Undercounts." *Social Science History* 19: 455-66.

King, Nigel and Christine Horrocks. 2010. *Interviews in Qualitative Research.* Thousand Oaks, CA: Sage.

Kirk, Andy. 2016. *Data Visualisation: A Handbook for Data Driven Design.* Thousand Oaks, CA: Sage.

Kirk, David S. and Andrew V. Papachristos. 2011. "Neighborhood Violence 1." *American Journal of Sociology* 116(4):1190-233.

Kiser, Edgar and Michael Hechter. 1991. "The Role of General Theory in Comparative-Historical Sociology." *American Journal of Sociology* 97:1-30.

Kitchener, Karen Strohm and Richard F. Kitchener. 2009. "Social Science Research Ethics: Historical and Philosophical Issues." Pp. 5-22 in *The Handbook of Social Research Ethics,* edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Kitchin, Rob, Tracey P. Lauriault, and Matthew W. Wilson. 2017. *Understanding Spatial Media.* Thousand Oaks, CA: Sage.

Klein, Malcolm W. 1971. *Street Gangs and Street Workers.* Englewood Cliffs, NJ: Prentice Hall.

Klinenberg, Eric. 2002. *Heat Wave: A Social Autopsy of Disaster in Chicago.* Chicago, IL: University of Chicago Press.

Kochel, Tammy Rinehart. 2019. "Explaining Racial Differences in Ferguson's Impact on Local Residents' Trust and Perceived Legitimacy: Policy Implications for Police." *Criminal Justice Policy Review* 30(3):374-405.

Koegel, Paul. 1987. *Ethnographic Perspectives on Homeless and Homeless Mentally Ill Women.* Washington, DC: Alcohol, Drug Abuse, and Mental Health Administration, Public Health Service, U.S. Department of Health and Human Services.

Kohn, Alfie. 2008. "Who's Cheating Whom?" *The Education Digest* 73:4-11.

Kohn, Melvin L. 1987. "Cross-National Research as an Analytic Strategy." *American Sociological Review* 52:713-31.

Kohut, Andrew, Scott Keeter, Carroll Doherty, Michael Dimock, and Leah Christian. 2012. *Assessing the Representativeness of Public Opinion Surveys.* Washington, DC: Pew Research Center (www.people-press.org/2012/05/15/assessingthe-representativeness-of-public-opinionsurveys).

Kohut, Andrew. 1988. "Polling: Does More Information Lead to Better Understanding?" *The Boston Globe,* November 7, 25.

Kolbert, Elizabeth. 1992. "Test-Marketing a President." *The New York Times Magazine,* August 30, 18-21, 60, 68, 72.

Kollock, Peter and Marc A. Smith. 1999. "Communities in Cyberspace." Pp. 3-25 in *Communities in Cyberspace,* edited by Peter Kollock and Marc A. Smith. New York, NY: Routledge.

Korn, James H. 1997. *Illusions of Reality: A History of Deception in Social Psychology.* Albany, NY: State University of New York Press.

Kotkin, Stephen. 2002. "A World War Among Professors." *The New York Times,* September 7, Arts pp. 1, 17.

Kozinets, Robert V. 2010. *Netnography: Doing Ethnographic Research Online.* Thousand Oaks, CA: Sage.

Kraemer, Helena Chmura and Sue Thiemann. 1987. *How Many Subjects? Statistical Power Analysis in Research.* Newbury Park, CA: Sage.

Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks." *Proceedings of the National Academy of Sciences (PNAS),* 111:8788-790.

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity." *Public Opinion Quarterly* 72:847-65.

Krieger, Lisa M. 2021. "The Fall From Grace for Johnson & Johnson's COVID-19 Vaccine." *The Mercury News,* May.

Kroenke, Kurt, Robert L. Spitzer, Janet B. W. Williams, and Bernd Löwe. 2009. "An Ultra-Brief Screening Scale for Anxiety and Depression: The PHQ-4." *Psychosomatics* 50(6):613-21.

Krogstad, Jens Manuel, Mark Hugo Lopez, and Molly Rohal. 2015. *English Proficiency on the Rise Among Latinos: U.S. Born Driving Language Changes.* Washington, DC: Pew Research Center (http://www.pewhispanic.org/2015/05/12/englishproficiency-on-the-rise-among-latinos/).

Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50:537-67.

Krosnick, Jon A. 2006. *The Handbook of Questionnaire Design: Insights from Social and Cognitive Psychology.* Eric M. Mindich Encounters With Authors Symposium at Institute for Quantitative Social Science, Harvard University, Cambridge MA, January 19–21.

Krueger, Richard A. and Mary Anne Casey. 2009. *Focus Groups: A Practical Guide for Applied Research,* 4th ed. Thousand Oaks, CA: Sage.

Kubey, Robert. 1990. "Television and the Quality of Family Life." *Communication Quarterly* 38(Fall):312-24.

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions,* 2nd ed. Chicago, IL: University of Chicago Press.

Kuzel, Anton J. 1999. "Sampling in Qualitative Inquiry." Pp. 33-45 in *Doing Qualitative Research,* 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Kvale, Steinar. 1996. *Interviews: An Introduction to Qualitative Research Interviewing.* Thousand Oaks, CA: Sage.

Kvale, Steinar. 2002. "The Social Construction of Validity." Pp. 299-325 in *The Qualitative Inquiry Reader,* edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Labaw, Patricia J. 1980. *Advanced Questionnaire Design.* Cambridge, MA: ABT Books.

Lacey, John H., Tara Kelley-Baker, Robert B. Voas, Eduardo Romano, C. Debra Furr-Holden, Pedro Torres, and Amy Berning. 2011. "Alcohol- and Drug-Involved Driving in the United States: Methodology for the 2007 National Roadside Survey." *Evaluation Review* 35:319-53.

Lahman, Maria K. E. 2018. *Ethics in Social Research: Becoming Culturally Responsive.* Thousand Oaks, CA: Sage.

Lange, Matthew. 2013. *Comparative–Historical Methods.* Thousand Oaks, CA: Sage.

Langford, Terri. 2000. "Census Workers in Dallas Find the Well-Off Hard to Count." *The Boston Globe,* June 1, A24.

LaRossa, Ralph. 1995. *Parenthood in Early Twentieth-Century America Project (PETCAP), 1900–1944* [Computer file]. Atlanta, GA: Georgia State University [producer], 1995. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1997.

Larson, Calvin J. 1993. *Pure and Applied Sociological Theory: Problems and Issues.* New York, NY: Harcourt Brace Jovanovich.

Lathrop, Barnes F. 1968. "History From the Census Returns." Pp. 79-101 in *Sociology and History: Methods,* edited by Seymour Martin Lipset and Richard Hofstadter. New York, NY: Basic Books.

Latour, Francie. 2002. "Marching Orders: After 10 Years, State Closes Prison Boot Camp." *Boston Sunday Globe,* June 16, B1, B7.

Lavin, Michael R. 1994. *Understanding the 1990 Census: A Guide for Marketers, Planners, Grant Writers and Other Data Users.* Kenmore, NY: Epoch Books.

Lavrakas, Paul J. 1987. *Telephone Survey Methods: Sampling, Selection, and Supervision.* New-Bury Park, CA: Sage.

Lavrakas, Paul J., Grant Benson, Stephen Blumberg, Trent Buskirk, Ismael Flores Cervantes, Leah Christian et al. 2017. Report From the AAPOR Task Force on "The Future of U.S. General Population Telephone Survey Research" (http://www.aapor.org/getattachment/Education-Resources/Reports/Future-of-Telephone-Survey-Research-Report.pdf.aspx).

Layte, Richard and Christopher T. Whelan. 2003. "Moving in and Out of Poverty: The Impact of Welfare Regimes on Poverty Dynamics in the EU." *European Societies* 5:167-91.

Lazer, David and Jason Radford. 2017. "Data ex Machina: Introduction to Big Data." *Annual Review of Sociology* 43:19-39.

Leakey, Tricia, Kevin B. Lunde, Karin Koga, and Karen Glanz. 2004. "Written Parental Consent and the Use of Incentives in a Youth Smoking Prevention Trial: A Case Study From Project SPLASH." *American Journal of Evaluation* 25:509-23.

LeDuc, Lawrence, Richard G. Niemi, and Pippa Norris (Eds.). 1996. *Comparing Democracies: Elections and Voting in Global Perspective.* Thousand Oaks, CA: Sage.

Legewie, Joscha and Jeffrey Fagan. 2019. "Aggressive Policing and the Educational Performance of Minority Youth." *American Sociological Review* 84(2):220-47.

Legewie, Joscha. 2016. "Racial Profiling and Use of Force in Police Stops: How Local Events Trigger Periods of Increased Discrimination." *American Journal of Sociology* 122(2):379-424.

Lehrer, Jonah. 2010. "The Truth Wears Off: Is There Something Wrong With the Scientific Method?" *The New Yorker,* December 13:52-57.

Lempert, Richard. 1989. "Humility Is a Virtue: On the Publicization of Policy-Relevant Research." *Law & Society Review* 23:146-61.

Lempert, Richard and Joseph Sanders. 1986. *An Invitation to Law and Social Science: Desert, Disputes, and Distribution.* New York, NY: Longman.

Leonhardt, David. 2020. "Wrong Again How Polls Misread 2020 Voters." *The New York Times,* A1.

Leonhardt, David. 2021. "New Study Elucidates the Benefits of Pre-K." *The New York Times,* May 10. Published online as "The Power of Pre-K."

Levine, James P. 1976. "The Potential for Crime Overreporting in Criminal Victimization Surveys." *Criminology* 14:307-30.

Levine, Jeremy R. 2021. *Constructing Community: Urban Governance, Development, and Inequality in Boston.* Princeton, NJ: Princeton University Press.

Levinson, Martin P. 2010. "Accountability to Research Participants: Unresolved Dilemmas and Unravelling Ethics." *Ethnography and Education* 5:193-207.

Levitas, Ruth and Will Guy. 1996. "Introduction." Pp. 1-6 in *Interpreting Official Statistics,* edited by Ruth Levitas and Will Guy. New York, NY: Routledge.

Levitt, Heidi M., Rebecca Todd Swanger, and Jenny B. Butler. 2008. "Male Perpetrators' Perspectives on Intimate Partner Violence, Religion, and Masculinity." *Sex Roles* 58:435-48.

Levy, Paul S. and Stanley Lemeshow. 1999. *Sampling of Populations: Methods and Applications,* 3rd ed. New York, NY: Wiley.

Lewin, Tamar. 2001a. "Income Education Is Found to Lower Risk of New Arrest." *The New York Times,* November 16, A18.

Lewin, Tamar. 2001b. "Surprising Result in Welfare-to-Work Studies." *The New York Times,* July 31, A16.

Lewis, Kevin, Nicholas Christakis, Marco Gonzalez, Jason Kaufman, and Andreas Wimmer. 2008. "Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com." *Social Networks* 30(4):330-42.

Lewis, Kevin. 2015. "Three Fallacies of Digital Footprints." *Big Data & Society* (July-December):1-4.

Lewis-Beck, Michael S., Alan Bryman, and Tim Futing Liao (Eds.). 2004. *The SAGE Encyclopedia of Social Science Research Methods,* vol. 1. Thousand Oaks, CA: Sage.

Li, Jun. 2008. "Ethical Challenges in Participant Observation: A Reflection on Ethnographic Fieldwork." *The Qualitative Report* 13:100-15.

Liberatore, Stacy. 2016. "What Happens in an Internet Second: 54,907 Google Searches, 7,252 Tweets, 125,406 YouTube Video Views and 2,501,018 Emails Sent." *Daily Mail* (June 27) (http://www.dailymail.co.uk/sciencetech/article-3662925/What-happens-internet-second-54-907Google-searches-7-252-tweets-125-406-YouTube-video-views-2-501-018-emails-sent.html).

Lieberson, Stanley. 1985. *Making It Count: The Improvement of Social Research and Theory.* Berkeley, CA: University of California Press.

Lieberson, Stanley. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70:307-20.

Likert, Rensis. 1932. "A Technique for Measurement of Attitudes." *Archives of Psychology* 140:5-55.

Lillard, Lee A. and Constantijn W. A. Panis. 1998. "Panel Attrition From the Panel Study of Income Dynamics: Household Income, Marital Status, and Mortality." *The Journal of Human Resources* 33:437-57.

Lincoln, Yvonna S. 2009. "Ethical Practices in Qualitative Research." Pp. 150-69 in *The Handbook of Social Research Ethics,* edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Lindsay, Sally, Simon Smith, Frances Bell, and Paul Bellaby. 2007. "Tackling the Digital Divide: Exploring the Impact of ICT on Managing Heart Conditions in a Deprived Area." *Information, Communication & Society* 10:95-114.

Ling, Rich and Gitte Stald. 2010. "Mobile Communities: Are We Talking About a Village, a Clan, or a Small Group?" *American Behavioral Scientist* 53(8):113-1147.

Link, Bruce G., Jo C. Phelan, Ann Stueve, Robert E. Moore, Michaeline Brenahan, and Elmer L. Struening. 1996. "Public Attitudes and Beliefs About Homeless People." Pp. 143-48 in *Homelessness in America,* edited by Jim Baumohl. Phoenix, AZ: Oryx.

Lipset, Seymour Martin. 1968. *Revolution and Counterrevolution.* New York, NY: Basic Books.

Lipset, Seymour Martin, Martin Trow, and James Coleman. 1956. *Union Democracy: The Internal Politics of the International Typographical Union.* New York, NY: Doubleday Anchor.

Lipsey, Mark W. and David B. Wilson. 2001. *Practical Meta-Analysis.* Thousand Oaks, CA: Sage.

Lipsky, Michael. 1980. *Street-Level Bureaucracy.* New York, NY: Russell Sage Foundation.

Liptak, Kevin. 2012. "Report Shows Turnout Lower Than 2008 and 2004." CNN (http://politicalticker.blogs.cnn.com/2012/11/08/report-shows-turnoutlower-than-2008-and-2004).

Litwin, Mark S. 1995. *How to Measure Survey Reliability and Validity.* Thousand Oaks, CA: Sage.

Loader, Brian D., Steve Muncer, Roger Burrows, Nicolas Pleace, and Sarah Nettleton. 2002. "Medicine on the Line? Computer-Mediated Social Support and Advice for People With Diabetes." *International Journal of Social Welfare* 11:53-65.

Locke, Lawrence F., Stephen J. Silverman, and Waneen Wyrick Spirduso. 1998. *Reading and Understanding Research.* Thousand Oaks, CA: Sage.

Locke, Lawrence F., Waneen Wyrick Spirduso, and Stephen J. Silverman. 2000. *Proposals That Work: A Guide for Planning Dissertations and Grant Proposals,* 4th ed. Thousand Oaks, CA: Sage.

Lockwood, Daniel. 1996. *Violent Incidents Among Selected Public School Students in Two Large Cities of the South and the Southern Midwest, 1995:* [United States] [Computer file]. ICPSR version. Atlanta, GA: Clark Atlantic University [producer], 1996. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1998.

Lofland, John and Lyn H. Lofland. 1984. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis,* 2nd ed. Belmont, CA: Wadsworth.

Lund, Laura and William E. Wright. 1994. "Mitofsky–Waksberg vs. Screened Random Digit Dial: Report on a Comparison of the Sample Characteristics of Two RDD Survey Designs." Presented at the Center for Disease Control's 11th Annual Behavioral Risk Factor Survey Conference, Atlanta, GA, June.

Luxardo, Natalia, Graciela Colombo, and Gabriela Iglesias. 2011. "Methodological and Ethical Dilemmas Encountered During Field Research of Family Violence Experienced by Adolescent Women in Buenos Aires." *The Qualitative Report* 16:984-1000.

Lyall, Sarah. 2004. "Does Queen Get Her Mail on Time? You've Got to Wonder." *The New York Times,* May 28, A4.

Lyberg, Lars, and Patricia Dean. 1992. "Methods for Reducing Nonresponse Rates: A Review." Annual Meeting of the American Association for Public Opinion Research. St. Petersburg, Florida.

Lynch, Michael and David Bogen. 1997. "Sociology's Asociological 'Core': An Examination of Textbook Sociology in Light of the Sociology of Scientific Knowledge." *American Sociological Review* 62:481-93.

Mabry, Linda. 2008. "Case Study in Social Research." Pp. 214-27 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

MacDougall, Colin and Frances Baum. 1997. "The Devil's Advocate: A Strategy to Avoid Groupthink and Stimulate Discussion in Focus Groups." *Qualitative Health Research* 7:532-41.

MacFarquhar, Neil. 2021a. "With Homicides Rising, Cities Brace for a Violent Summer." *The New York Times,* A14.

MacFarquhar, Neil. 2021b. "Murders Spike at Record Rate Across the U.S." *The New York Times,* A1, A19.

Madden, Raymond. 2010. *Being Ethnographic: A Guide to the Theory and Practice of Ethnography.* Thousand Oaks, CA: Sage.

Magson, Natasha R., Justin Y. A. Freeman, Ronald M. Rapee, Cele E. Richardson, Ella L. Oar, and Jasmine Fardouly. 2021. "Risk and Protective Factors for Prospective Changes in Adolescent Mental Health during the COVID-19 Pandemic." *Journal of Youth and Adolescence* 50(1):44-57.

Mahoney, James. 2001. *The Legacies of Liberalism: Path Dependence and Political Regimes in Central America.* New York, NY: Cambridge University Press.

Majchrzak, Ann and M. Lynne Markus. 2014. *Methods for Policy Research: Taking Socially Responsible Action,* 2nd ed. Thousand Oaks, CA: Sage.

Mandel, Benjamin A. and Mark S. Ludwick. 2021. *COVID-19 Pandemic: Federal Recovery Legislation and the NIPAs.* Washington DC: U.S. Department of Commerce, Bureau of Economic Analysis.

Mangione, Thomas W. 1995. *Mail Surveys: Improving the Quality.* Thousand Oaks, CA: Sage.

Manza, Jeff, Clem Brooks, and Michael Sauder. 2005. "Money, Participation, and Votes: Social Cleavages and Electoral Politics." Pp. 201-26 in *The Handbook of Political Sociology: States, Civil Societies, and Globalization,* edited by Thomas Janoski, Robert R. Alford, Alexander M. Hicks, and Mildred A. Schwartz. New York, NY: Cambridge University Press.

Margolis, Eric. 2004. "Looking at Discipline, Looking at Labour: Photographic Representations of Indian Boarding Schools." *Visual Studies* 19:72-96.

Marini, Margaret Mooney and Burton Singer. 1988. "Causality in the Social Sciences." Pp. 347-409 in *Sociological Methodology,* vol. 18, edited by Clifford C. Clogg. Washington, DC: American Sociological Association.

Mark, Melvin M. and Chris Gamble. 2009. "Experiments, Quasi-Experiments, and Ethics." Pp. 198-213 in *The Handbook of Social Research Ethics,* edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Markham, Annette N. 2008. "The Methods, Politics, and Ethics of Representation in Online Ethnography." Pp. 247-84 in *Collecting and Interpreting Qualitative Materials,* 3rd ed., edited by Norman Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Markoff, John. 2005. "Transitions to Democracy." Pp. 384-403 in *The Handbook of Political Sociology: States, Civil Societies, and Globalization,* edited by Thomas Janoski, Robert R. Alford, Alexander M. Hicks, and Mildred A. Schwartz. New York, NY: Cambridge University Press.

Marozzi, Marco. 2016. "Construction, Robustness Assessment and Application of an Index of Perceived Level of Socio-Economic Threat from Immigrants: A Study of 47 European Countries and Regions." *Social Indicators Research* 128(1):413-37.

Marsden, Peter V. 1987. "Core Discussion Networks of Americans." *American Sociological Review* 52:122-31.

Marsden, Peter V. and Tom W. Smith. 2016. *Overview: The General Social Survey Project.* GSS Project Report No. 31. Chicago, IL: National Opinion Research Center (NORC).

Marshall, Gary D. and Philip G. Zimbardo. 1979. "Affective Consequences of Inadequately Explained Physiological Arousal." *Journal of Personality and Social Psychology* 37:970-88.

Martin, Lawrence L. and Peter M. Kettner. 1996. *Measuring the Performance of Human Service Programs.* Thousand Oaks, CA: Sage.

Martin, Linda G. and Kevin Kinsella. 1995. "Research on the Demography of Aging in Developing Countries." Pp. 356-403 in *Demography of Aging,* edited by Linda G. Martin and Samuel H. Preston. Washington, DC: National Academy Press.

Martínez, Daniel E., Jeremy Slack, Kraig Beyerlein, Prescott Vandervoet, Kristin Klingman, Paola Molina, Shiras Manning, Melissa Burham, Kylie Walzak, Kristen Valencia, and Lorenzo Gamboa. 2017. "The Migrant Border Crossing Study: A Methodological Overview of Research along the Sonora–Arizona Border." *Population Studies* 71(2):249-64.

Marx, Karl. 1967. *Capital: A Critique of Political Economy.* New York, NY: International Publishers.

Marx, Karl and Friedrich Engels. 1961. "The Communist Manifesto." Pp. 13-44 in *Essential Works of Marxism,* edited by Arthur P. Men-del. New York, NY: Bantam.

Matejka, Justin, Michael Glueck, Tovi Grossman and George Fitzmaurice. 2016. "The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales." *CHI* (http://www.dgp.toronto.edu/~tovi/papers/vas.pdf).

Mathews, Kelly, Jessica Phelan, Nicholas A. Jones, Sarah Konya, Rachel Marks, Beverly M. Pratt, Julia Coombs, and Michael Bentley. 2017. *2015 National Content Test: Race and Ethnicity Analysis Report.* Washington DC: U.S. Department of Commerce, Census Bureau.

Matt, Georg E. and Thomas D. Cook. 1994. "Threats to the Validity of Research Syntheses." Pp. 503-20 in *The Handbook of Research Synthesis,* edited by Harris Cooper and Larry V. Hedges. New York, NY: Russell Sage Foundation.

Maxwell, Joseph A. 2005. *Qualitative Research Design: An Interactive Approach,* 2nd ed. Thousand Oaks, CA: Sage.

Mayer-Schönberger, Viktor and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think.* Boston, MA: Houghton Mifflin Harcourt.

Maynard, Douglas W., Jeremy Freese, and Nora Cate Schaeffer. 2010. "Calling for Participation: Requests, Blocking Moves, and Rational (Inter)action in Survey Introductions." *American Sociological Review* 75:791-814.

Mayrl, Damon, Ben Moodie, Jon Norman, Jodi Short, Sarah Staveteig, and Cinzia Solari. 2004. "A Theory of Relativity." *Contexts* 3:10.

McCarter, Susan A. 2009. "Legal and Extralegal Factors Affecting Minority Overrepresentation in Virginia's Juvenile Justice System: A Mixed-Method Study." *Child and Adolescent Social Work Journal* 26:533-44.

McConahay, John B. 1986. "Modern Racism, Ambivalence, and the Modern Racism Scale." Pp. 91-125 in *Prejudice, Discrimination, and Racism,* edited by John F. Davidio and Samuel L. Gaertner. New York, NY: Academic Press.

McGeeney, Kyley and Scott Keeter. 2014. "Pew Research Increases Share of Interviews Conducted by Cellphone." Pew Research Center, January 15 (www.pewresearch.org/fact-tank/2014/01/15/pewresearch-increases-share-of-interviewsconducted-by-cellphone).

McIntyre, Alice. 2008. *Participatory Action Research.* Thousand Oaks, CA: Sage.

McLellan, A. Thomas, Lester Luborsky, John Cacciola, Jeffrey Griffith, Frederick Evans, Harriet L. Barr, and Charles P. O'Brien. 1985. "New Data F rom the Addiction Severity Index: Reliability and Validity in Three Centers." *The Journal of Nervous and Mental Disease* 173(7):412-23.

McPhail, Clark and John McCarthy. 2004. "Who Counts and How: Estimating the Size of Protests." *Contexts* 3:12-18.

McPherson, Miller, Lynn Smith-Lovin, and Matthew E. Brashears. 2006. "Social Isolation in America: Changes in Core Discussion Networks Over Two Decades." *American Sociological Review* 71:353-75.

Medway, Rebecca L. and Jenna Fulton. 2012. "When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates." *Public Opinion Quarterly* 76:733-46.

Meghani, S. H., Brooks, J., Gipson-Jones, T., Waite, R., Whitefield-Harris, L., Deatrick, J. 2009. "Patient–Provider Race-Concordance: Does It Matter in Improving Minority Patients' Health Outcomes?" *Ethnicity & Health* 14(1):107-30.

Menikoff, Jerry, Julie Kaneshiro, and Ivor Pritchard. 2017. "The Common Rule, Updated." *The New England Journal of Medicine* 376(7):613-15.

Mertens, Donna M. 2012. "Transformative Mixed Methods: Addressing Inequities." *American Behavioral Scientist* 56:802-13.

Merton, Robert K., Marjorie Fiske, and Patricia L. Kendall. 1956. *The Focused Interview.* Glencoe, IL: Free Press.

Michel, Franck. 2012. "How Many Photos Are Uploaded to Flickr Every Day, Month, Year." Flickr, March 20 (https://www.flickr.com/photos/franckmichel/6855169886).

Mieczkowski, Tom. 1997. "Hair Assays and Urinalysis Results for Juvenile Drug Offenders." *National Institute of Justice Research Preview.* Washington, DC: U.S. Department of Justice.

Milbrath, Lester and M. L. Goel. 1977. *Political Participation,* 2nd ed. Chicago, IL: Rand McNally.

Miles, Matthew B. and A. Michael Huberman. 1994. *Qualitative Data Analysis,* 2nd ed. Thousand Oaks, CA: Sage.

Miles, Matthew B., A. Michael Huberman, and Johnny Soldaña. 2014. *Qualitative Data Analysis: A Methods Sourcebook,* 3rd ed. Thousand Oaks: Sage.

Milgram, Stanley. 1963. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology* 67:371-478.

Milgram, Stanley. 1964. "Issues in the Study of Obedience: A Reply to Baumrind." *American Psychologist* 19:848-52.

Milgram, Stanley. 1965. "Some Conditions of Obedience and Disobedience to Authority." *Human Relations* 18:57-76.

Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View.* New York, NY: Harper & Row.

Milgram, Stanley. 1977. "Subject Reaction: The Neglected Factor in the Ethics of Experimentation." *Hastings Law Review,* October, pp. 19-23 (as cited in Cave, Emma and Soren Holm. 2003. "Milgram and Tuskegee—Paradigm Research Projects in Bioethics." *Health Care Analysis* 11:27-40).

Milgram, Stanley. 1992. *The Individual in a Social World,* 2nd ed. New York, NY: McGraw-Hill.

Mill, John Stuart. 1872. *A System of Logic: Ratiocinative and Inductive,* 8th ed., vol. 2. London: Longmans, Green, Reader, & Dyer.

Miller, Arthur G. 1986. *The Obedience Experiments: A Case Study of Controversy in Social Science.* New York, NY: Praeger.

Miller, Delbert C. and Neil J. Salkind. 2002. *Handbook of Research Design and Social Measurement,* 6th ed. Thousand Oaks, CA: Sage.

Miller, JoAnn. 2003. "An Arresting Experiment: Domestic Violence Victim Experiences and Perceptions." *Journal of Interpersonal Violence* 18:695-716.

Miller, Sara G. and Reuters. 2021. "Johnson & Johnson Vaccine Linked to 28 Cases of Blood Clots, CDC Reports." *Nbcnews.Com,* May 12.

Miller, Stuart S., Conor J. O'Dea, and Donald A. Saucier. 2021. "'I Can't Breathe': Lay Conceptualizations of Racism Predict Support for Black Lives Matter." *Personality and Individual Differences* 173(December 2020):110625.

Miller, Susan. 1999. *Gender and Community Policing: Walking the Talk.* Boston, MA: Northeastern University Press.

Miller, Ted R. and Delia Hendrie. 2008. *Substance Abuse Prevention Dollars and Cents: A Cost-Benefit Analysis,* DHHS Pub. No. (SMA) 07-4298. Rockville, MD: Center for Substance Abuse Prevention, Substance Abuse and Mental Health Services Administration.

Miller, Walter B. 1992. *Crime by Youth Gangs and Groups in the United States.* Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Miller, William L. and Benjamin F. Crabtree. 1999a. "Clinical Research: A Multimethod Typology and Qualitative Roadmap." Pp. 3-30 in *Doing Qualitative Research,* 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Miller, William L. and Benjamin F. Crabtree. 1999b. "The Dance of Interpretation." Pp. 127-43 in *Doing Qualitative Research,* edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Miller, William L. and Benjamin F. Crabtree. 1999c. "Depth Interviewing." Pp. 89-107 in *Doing Qualitative Research,* 2nd ed., edited by Benjamin F. Crabtree and William L. Miller. Thousand Oaks, CA: Sage.

Mills, C. Wright. 1959. *The Sociological Imagination.* New York, NY: Oxford University Press.

Minkler, Meredith. 2000. "Using Participatory Action Research to Build Healthy Communities." *Public Health Reports* 115:191-97.

Mirowsky, John. 1999. *Aging, Status, and the Sense of Control: Competing Continuation* (Proposal to the National Institute of Aging). Urbana, IL: University of Illinois Press.

Mirowsky, John and Catherine E. Ross. 1999. "Economic Hardship Across the Life Course." *American Sociological Review* 64:548-69.

Mirowsky, John and Catherine E. Ross. 2001. *Aging, Status, and the Sense of Control (ASOC), 1995, 1998, 2001 [United States] Questionnaire* (ICPSR 3334). Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Mirowsky, John and Catherine E. Ross. 2003. *Education, Social Status, and Health.* New York, NY: Aldine de Gruyter.

Mitchell, Richard G. Jr. 1993. *Secrecy and Fieldwork.* Newbury Park, CA: Sage.

Moallem, J. 2020. "This is How You Live When the World Fall Apart." *The New Yok Times,* March 12, SR4.

Mobasseri, Sanaz. 2019. *Race, Place, and Crime: How Violent Crime Events Affect Employment Discrimination.* Chicago, IL: University of Chicago Press.

Moe, Angela M. 2007. "Silenced Voices and Structural Survival—Battered Women's Help Seeking." *Violence Against Women* 13(7):676-99.

Moe, Hallvard and Anders Olof Larsson. 2012. "Methodological and Ethical Challenges Associated With Large-Scale Analyses of Online Political Communication." *NORDICOM Review: Nordic Research on Media and Communication* 33(1):117-24.

Mohr, Lawrence B. 1992. *Impact Analysis for Program Evaluation.* Newbury Park, CA: Sage.

Monkkonen, Eric H. 1994. "Introduction." Pp. 1-8 in *Engaging the Past: The Uses of History Across the Social Sciences.* Durham, NC: Duke University Press.

Mooney, Christopher Z. and Mei Hsien Lee. 1995. "Legislating Morality in the American States: The Case of Abortion Regulation Reform." *American Journal of Political Science* 39:599-627.

Moore, Barrington Jr. 1966. *Social Origins of Democracy and Dictatorship: Lord and Peasant in the Making of the Modern World.* Boston, MA: Beacon Press.

Moore, Spencer, Mark Daniel, Laura Linnan, Marci Campbell, Salli Benedict, and Andrea Meier. 2004. "After Hurricane Floyd Passed: Investigating the Social Determinants of Disaster Preparedness and Recovery." *Family and Community Health* 27:204-17.

Morgan, Anthony, Hayley Boxall, and Australian Institute of Criminology. n.d. *Social Isolation, Time Spent at Home, Financial Stress and Domestic Violence during the COVID-19 Pandemic.*

Morgan, David L. 2014. *Integrating Qualitative & Quantitative Methods: A Pragmatic Approach.* Thousand Oaks, CA: Sage.

Morgan, Stephen L. 2020. "Response Rates and Representativeness: A Benchmark Comparison of the General Social Surveys to the American Community Surveys, 2012-2018*."

Morrill, Calvin, Christine Yalda, Madeleine Adelman, Michael Musheno, and Cindy Bejarano. 2000. "Telling Tales in School: Youth Culture and Conflict Narratives." *Law & Society Review* 34:521–65.

Mosher, Clayton J., Terance D. Miethe, and Dretha M. Phillips. 2002. *The Mismeasure of Crime.* Thousand Oaks, CA: Sage.

Muhr, Thomas and Susanne Friese. 2004. *User's Manual for ATLAS.ti 5.0,* 2nd ed. Berlin: Scientific Software Development.

Myers, Steven Lee. 2002. "Russia Takes Stock of a Nation's Transformation." *The New York Times,* September 29, 3.

Myers, Steven Lee. 2010. "Delays in a Head Count Keep Crucial Numbers a Matter of Guesswork." *The New York Times,* December 7, A10.

Nadal, Kevin L., Kristin C. Davidoff, Neil Allicock, Christine R. Serpe, and Tanya Erazo. 2017. "Perceptions of Police, Racial Profiling, and Psychological Outcomes: A Mixed Methodological Study." *Journal of Social Issues* 73(4):808-30.

Nagourney, Adam. 2002. "Cellphones and Caller ID Are Making It Harder for Pollsters to Pick a Winner." *The New York Times,* November 5, A20.

Nakonezny, Paul A., Rebecca Reddick, and Joseph Lee Rodgers. 2004. "Did Divorces Decline After the Oklahoma City Bombing?" *Journal of Marriage and Family* 66:90-100.

Narayan, Sowmya and Jon A. Krosnick. 1996. "Education Moderates Some Response Effects in Attitude Measurement." *Public Opinion Quarterly* 60:58-88.

National Center for Bioethics in Research and Health Care. n.d. *About the USPHS Syphilis Study* (http://tuskegeebioethics.org/about-the-usphs-syphilis-study/).

National Geographic Society. 2000. *Survey 2000* (http://survey2000.nationalgeographic.com).

National Institute on Alcohol Abuse and Alcoholism (NIAAA). 1994. "Alcohol-Related Impairment." *Alcohol Alert* 25 (July):1–5.

National Institute on Alcohol Abuse and Alcoholism (NIAAA). 1997. "Alcohol Metabolism." *Alcohol Alert* 35(January):1–4.

National Institute on Alcohol Abuse and Alcoholism (NIAAA). 2013. *NIAAA Recognizes Alcohol Awareness Month.* Retrieved from (www.niaaa.nih.gov/news-events/alcohol-awareness-month-2013).

National Institute on Alcohol Abuse and Alcoholism (NIAAA). n.d. "Alcohol Use Disorder." Retrieved from (https://www.niaaa.nih.gov/alcoholhealth/overview-alcohol-consumption/alcohol-use-disorders).

National Research Center (NORC). 2021. General Social Survey. Chicago: NORC at the University of Chicago (https://gss.norc.org/?msclkid=939d61d6bb4911ec9bf4091d3bcfb365).

National Oceanic and Atmospheric Administration (NOAA). 2005. *Hurricane Katrina.* Washington, DC: U.S. Department of Commerce (www.ncdc.noaa.gov/extremeevents/specialreports/Hurricane-Katrina.pdf).

National Oceanic and Atmospheric Administration (NOAA). 2017. *Hurricane Harvey.* Retrieved from (http://www.nhc.noaa.gov/archive/2017/HARVEY.shtml).

National Oceanic and Atmospheric Administration [NOAA]. 2021a. "Summer 2021 Neck and Neck With Dust Bowl Summer for Hottest on Record." Retrieved November 19, 2021 (https://www.noaa.gov/news/summer-2021-neck-and-neck-with-dust-bowl-summer-for-hottest-on-record).

National Oceanic and Atmospheric Administration [NOAA]. 2021b. "U.S. Hit With 18 Billion-Dollar Disasters So Far This Year." Retrieved November 19, 2021 (https://www.noaa.gov/news/us-hit-with-18-billion-dollar-disasters-so-far-year).

National Opinion Research Center (NORC). 2011. *General Social Survey (GSS).* Chicago, IL: NORC at the University of Chicago (www.norc.org/Research/Projects/Pages/general-social-survey.aspx).

National Opinion Research Center (NORC). 2014. General Social Survey (www3.norc.org/GSS+Website/Publications/GSS+Questionnaires).

National Opinion Research Center (NORC). 2016. Get the Data (http://gss.norc.org/Get-The-Data).

National Opinion Research Center (NORC). 2020. *General Social Survey.* Chicago, IL: National Opinion Research Center, University of Chicago.

National Opinion Research Center (NORC). 2021. *The General Social Survey (GSS).* Chicago, IL: NORC. Get GSS Data | NORC.

Needleman, Carolyn. 1981. "Discrepant Assumptions in Empirical Research: The Case of Juvenile Court Screening." *Social Problems* 28 (February):247-62.

Neil, Roland and Robert J. Sampson. 2021. "The Birth Lottery of History: Arrest Over the Life Course of Multiple Cohorts Coming of Age, 1995-2018." *American Journal of Sociology* 126(5):1127-78.

Neuendorf, Kimberly A. 2002. *The Content Analysis Guidebook.* Thousand Oaks, CA: Sage.

Newbury, Darren. 2005. "Editorial: The Challenge of Visual Studies." *Visual Studies* 20:1-3.

Newport, Frank. 1992. "Look at Polls as a Fever Chart of the Electorate." Letter to the Editor, *The New York Times,* November 6, A28.

Newport, Frank. 2016. "National Polling Accurately Nails Popular Vote." *Gallup,* November 23 (https://news.gallup.com/opinion/polling-matters/198155/national-polling-accurately-nails-popular-vote.aspx).

Nie, Norman H. and Lutz Erbring. 2002. *Internet and Society: A Preliminary Report.* Palo Alto, CA: Stanford Institute for the Quantitative Study of Society.

Nordanger, Dag. 2007. "Discourses of Loss and Bereavement in Tigray, Ethiopia." *Culture, Medicine and Psychiatry* 31:173-94.

Norris, Pippa. 2004. "The Bridging and Bonding Role of Online Communities." Pp. 31-41 in *Society Online: The Internet in Context,* edited by Philip N. Howard and Steve Jones. Thousand Oaks, CA: Sage.

Nunberg, Geoffrey. 2002. "The Shifting Lexicon of Race." *The New York Times,* December 22, WK3.

Nyhan, Brendan. 2021. "Why the Backfire Effect Does Not Explain the Durability of Political Misperceptions." *Proceedings of the National Academy of Sciences of the United States of America* 118(15).

Ó Dochartaigh, Niall. 2002. *The Internet Research Handbook: A Practical Guide for Students and Researchers in the Social Sciences.* Thousand Oaks, CA: Sage.

O'Brien, Daniel Tumminelli and Christopher Winship. 2017. "The Gains of Greater Granularity: The Presence and Persistence of Problem Properties in Urban Neighborhoods." *Journal of Quantitative Criminology* 33:649-74.

Office of Inspector General, U.S. Postal Service. 2017. *Delivery and Customer Service Operations–Bronx, NY* (https://www.uspsoig.gov/sites/default/files/documentlibrary-files/2017/DR-AR-17-004.pdf).

Ogburn, William F. 1930. "The Folkways of a Scientific Sociology," *Scientific Monthly* 30:300-06.

Olansky, Sidney, Stanley H. Schuman, Jesse J. Peters, C. A. Smith, Dorothy S. Rambo, and D. Washington. 1956. "Untreated Syphilis in the Male Negro X. Twenty Years of Clinical Observation of Untreated Syphilitic and Presumably Nonsyphilitic Groups." *Journal of Chronic Diseases* 4:177-85.

Olson, Kristen, Jolene D. Smyth, and Heather M. Wood. 2012. "Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation Rates? An Experimental Examination." *Public Opinion Quarterly* 76:611-35.

Olzak, Susan. 2021. "Does Protest Against Police Violence Matter? Evidence from U.S. Cities, 1990 through 2019." *American Sociological Review* 86(6):1066-99.

Ondersma, Steven J., Joanne Martin, Beverly Fortson, Daniel J. Whitaker, Shannon Self-Brown, Jessica Beatty, Amy Loree, David Bard, and Mark Chaffin. 2017. "Technology to Augment Early Home Visitation for Child Maltreatment Prevention: A Pragmatic Randomized Trial." *Child Maltreatment* (doi: 10.1177/1077559517729890).

Onoye, Jane M., Deborah A. Goebert, and Stephanie T. Nishimura. 2012. "Use of Incentives and Web-Based Administration for Surveying Student Alcohol and Substance Use in an Ethnically Diverse Sample." *Journal of Substance Use* 17(1):61-71.

Orcutt, James D. and J. Blake Turner. 1993. "Shocking Numbers and Graphic Accounts: Quantified Images of Drug Problems in the Print Media." *Social Problems* 49(May):190-206.

Organisation for Economic Co-operation and Development. 2017. "SF3.1: Marriage and Divorce Rates" (https://www.oecd.org/els/family/SF_3_1_Marriage_and_divorce_rates.pdf).

Orr, Larry L. 1999. *Social Experiments: Evaluating Public Programs With Experimental Methods.* Thousand Oaks, CA: Sage.

Orshansky, Mollie. 1977. "Memorandum for Daniel P. Moynihan. Subject: History of the Poverty Line." Pp. 232-37 in *The Measure of Poverty. Technical Paper I: Documentation of Background Information and Rationale for Current Poverty Matrix,* edited by Mollie Orshansky. Washington, DC: U.S. Department of Health, Education, and Welfare.

Orwin, Robert G. 1994. "Evaluating Coding Decisions." Pp. 138-62 in *The Handbook of Research Synthesis,* edited by Harris Cooper and Larry V. Hodges. New York, NY: Russell Sage Foundation.

Oude Groeniger, Joost, Kjell Noordzij, Jeroen van der Waal, and Willem de Koster. 2021. "Dutch COVID-19 Lockdown Measures Increased Trust in Government and Trust in Science: A Difference-in-Differences Analysis." *Social Science and Medicine* 275(February):113819.

Ousey, Graham C. and Matthew R. Lee. 2004. "Investigating the Connections Between Race, Illicit Drug Markets, and Legal Violence, 1984–1997." *Journal of Research in Crime and Delinquency* 41:352-83.

Pagnini, Deanna L. and S. Philip Morgan. 1996. "Racial Differences in Marriage and Childbearing: Oral History Evidence From the South in the Early Twentieth Century." *American Journal of Sociology* 101:1694-715.

Paige, Jeffery M. 1999. "Conjuncture, Comparison, and Conditional Theory in Macrosocial Inquiry." *American Journal of Sociology* 105:781-800.

Paika, Anthony and Kenneth Sanchagrina. 2013. "Social Isolation in America: An Artifact." *American Sociological Review* 78(3):339-60.

Pan, Yuling and Manuel de la Puente. 2005. *Census Bureau Guideline for the Translation of Data Collection Instruments and Supporting Materials: Documentation on How the Guideline Was Developed (Research Report Series, Survey Methodology #2005-06).* Washington, DC: Statistical Research Division, U.S. Census Bureau.

Panagopoulos, Costas. 2008. "Poll Accuracy in the 2008 Presidential Election" (www.fordham.edu/images/academics/graduate_schools/gsas/elections_and_campaign_/poll%20accuracy%20in%20the%202008%20presidential%20election.pdf).

Papineau, David. 1978. *For Science in the Social Sciences.* London: Macmillan.

Parker, Kim and Renee Stepler. 2017. "As U.S. Marriage Rate Hovers at 50%, Education Gap in Marital Status Widens." Pew Research Center (http://www.pewresearch.org/fact-tank/2017/09/14/as-u-s-marriagerate-hovers-at-50-education-gap-in-maritalstatus-widens/).

Parks, Kathleen A., Ann M. Pardi, and Clara M. Bradizza. 2006. "Collecting Data on Alcohol Use and Alcohol-Related Victimization: A Comparison of Telephone and Web-Based Survey Methods." *Journal of Studies on Alcohol* 67:318-23.

Parlett, Malcolm and David Hamilton. 1976. "Evaluation as Illumination: A New Approach to the Study of Innovative Programmes." Pp. 140-57 in *Evaluation Studies Review Annual,* vol. 1, edited by G. Glass. Beverly Hills, CA: Sage.

Pate, Antony M. and Edwin E. Hamilton. 1992. "Formal and Informal Deterrents to Domestic Violence: The Dade County Spouse Assault Experiment." *American Sociological Review* 57(October):691-97.

Paternoster, Raymond, Robert Brame, Ronet Bachman, and Lawrence W. Sherman. 1997. "Do Fair Procedures Matter? The Effect of Procedural Justice on Spouse Assault." *Law & Society Review* 31(1):163-204.

Patrick, Megan E. and Yvonne M. Terry-McElrath. 2017. "High Intensity Drinking by Underage Young Adults in the United States." *Addiction* 112:82-93.

Patterson, Orlando. 1997. "The Race Trap." *The New York Times,* July 11, A25.

Patton, Michael Quinn. 2002. *Qualitative Research & Evaluation Methods,* 3rd ed. Thousand Oaks, CA: Sage.

Patulny, Roger and Claire Seaman. 2017. "'I'll Just Text You': Is Face-to-Face Social Contact Declining in a Mediated World?" *Journal of Sociology* 53(2):285-302.

Pauwels, Luc. 2010. "Visual Sociology Reframed: An Analytical Synthesis and Discussion of Visual Methods in Social and Cultural Research." *Sociological Methods & Research* 38:545-81.

Paxton, Pamela. 2002. "Social Capital and Democracy: An Interdependent Relationship." *American Sociological Review* 67:254-77.

Pearson, Michael, Helen Sweeting, Patrick West, Robert Young, Jacki Gordon, and Katrina Turner. 2006. "Adolescent Substance Use in Different Social and Peer Contexts: A Social Network Analysis." *Drugs: Education, Prevention and Policy* 13:519-36.

Pedulla, David S. 2016. "Penalized or Protected? Gender and the Consequences of Nonstandard and Mismatched Employment Histories." *American Sociological Review* 81(2):262-89.

Peeples, Lynne. 2019. "What the Data Say About Police Shootings." *Nature* 573(7772):24-6.

Pepinsky, Harold E. 1980. "A Sociologist on Police Patrol." Pp. 223-34 in *Fieldwork Experience: Qualitative Approaches to Social Research,* edited by William B. Shaffir, Robert A. Stebbins, and Allan Turowetz. New York, NY: St. Martin's Press.

Percac-Lima, Sanja, Leslie S. Aldrich, Gloria B. Gamba, Adriana M. Bearse, and Steven J. Atlas. 2010. "Barriers to Follow-Up of an Abnormal Pap Smear in Latina Women Referred for Colposcopy." *Journal of General Internal Medicine* 25(11):1198-204.

Perrin, Andrew and Sara Atske. 2021. *7% of Americans Don't Use the Internet. Who Are They? Pew Research Center,* March 26.

Perrin, Andrew W. and Maeve Duggan. 2015. "Americans' Internet Access: 2000–2015." Pew Research Center, Internet & Technology, June 26 (http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/).

Perry, Gina. 2013. *Behind the Shock Machine: The Untold Story of the Notorious Milgram Psychology Experiments.* New York, NY: New Press.

Peters, Gerhard. "Election Year Presidential Preference Polling and Voting Outcomes." The American Presidency Project. John T. Woolley and Gerhard Peters (eds.). Santa Barbara, CA: University of California. Online. Downloaded 4/13/2022 from (https://www.presidency.ucsb.edu/node/323909).

Peterson, Robert A. 2000. *Constructing Effective Questionnaires.* Thousand Oaks, CA: Sage.

Petrosino, Anthony, Robert F. Boruch, Haluk Soydan, Lorna Duggan, and Julio Sanchez-Meca. 2001. "Meeting the Challenges of Evidence-Based Policy: The Campbell Collaboration." *Annals of the American Academy of Political and Social Science* 578(14):14-17.

Pew Hispanic Center. 2008. *Statistical Portrait of the Foreign-Born Population in the United States, 2006.* Statistical tables from 2000 Census and 2006 American Community Survey (http://pewhispanic.org/factsheets/factsheet.php?FactsheetID=36).

Pew Research Center. 2014. *Political Polarization in the American Public.* Washington, DC: Pew Research Center.

Pew Research Center. 2017a. *The American Trends Panel Survey Methodology* (http://assets.pewresearch.org/wp-content/uploads/sites/12/2017/09/14122154/W28-American-Trends-Panel-Survey-Methodology_CHECKED.pdf).

Pew Research Center. 2017b. *2017 Pew Research Center's American Trends Panel* (http://assets.pewresearch.org/wp-content/uploads/sites/12/2017/09/14122053/W28-Topline_marriage_blog-post_FINAL.pdf).

Pew Research Center. 2021. *Mobile Fact Sheet.* Washington, DC: Pew Research Center. Retrieved November 18, 2021 (https://www.pewresearch.org/internet/fact-sheet/mobile/).

Phillips, David P. 1982. "The Impact of Fictional Television Stories on U.S. Adult Fatalities: New Evidence on the Effect of the Mass Media on Violence." *American Journal of Sociology* 87:1340-59.

Philpot, Tasha S. and Kenneth M. Miller. 2020. "A New Face to the Race Card? Campaigns, Racial Cues, and Candidate Credibility." *Social Science Quarterly* 101(1):73-90.

Piliavin, Jane Allyn and Irving M. Piliavin. 1972. "Effect of Blood on Reactions to a Victim." *Journal of Personality and Social Psychology* 23:353-61.

Pink, Sarah. 2021. *Doing Visual Ethnography.* 4th ed. Thousand Oaks, CA: Sage.

Pink, Sarah, Heather Horst, John Postill, Larissa Hjorth, Tania Lewis, and Jo Tacchi. 2016. *Digital Ethnography: Principles and Practice.* Thousand Oaks, CA: Sage.

Pinterits, E. Janie, Poteat, V. Paul, & Spanierman, Lisa B. 2009. "The White Privilege Attitudes Scale: Development and Initial Validation." *Journal of Counseling Psychology* 56(3):417-29.

Plano Clark, Vicki L. and Nataliya V. Ivankova. 2016. *Mixed Methods Research: A Guide to the Field.* Thousand Oaks, CA: Sage.

Plessy v. Ferguson, 163 U.S. 537 (1896).

Pollner, Melvin and Richard E. Adams. 1994. "The Interpersonal Context of Mental Health Interviews." *Journal of Health and Social Behavior* 35:283-90.

Popper, Sir Karl. 1968/1934. *The Logic of Scientific Discovery.* New York, NY: Harper & Row.

Porter, Eduardo. 2021. "Study Shows Which Firms Discriminate." *The New York Times,* July 29.

Porter, Stephen R. and Michael E. Whitcomb. 2003. "The Impact of Contact Type on Web Survey Response Rates." *Public Opinion Quarterly* 67:579-88.

Posavac, Emil J. and Raymond G. Carey. 1997. *Program Evaluation: Methods and Case Studies,* 5th ed. Upper Saddle River, NJ: Prentice Hall.

Poth, Cheryl N. 2021. *Little Quick Fix: Research Ethics.* Thousand Oaks, CA: Sage

Presley, Cheryl A., Philip W. Meilman, and Rob Lyerla. 1994. "Development of the Core Alcohol and Drug Survey: Initial Findings and Future Directions." *Journal of American College Health* 42:248-55.

Presser, Stanley and Johnny Blair. 1994. "Survey Pretesting: Do Different Methods Produce Different Results?" *Sociological Methodology* 24:73-104.

Presser, Stanley, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, Jennifer M. Rothgeb, and Eleanor Singer. 2004. "Methods for Testing and Evaluating Survey Questions." *Public Opinion Quarterly* 68:109-30.

Price, Richard H., Michelle Van Ryn, and Amiram D. Vinokur. 1992. "Impact of a Preventive Job Search Intervention on the Likelihood of Depression Among the Unemployed." *Journal of Health and Social Behavior* 33 (June):158-67.

Przybylski, Liz. 2020. *Hybrid Ethnography.* New York, NY: Sage.

Puchta, Claudia and Jonathan Potter. 2004. *Focus Group Practice.* Thousand Oaks, CA: Sage.

Punch, Maurice. 1994. "Politics and Ethics in Qualitative Research." Pp. 83-97 in *Handbook of Qualitative Research,* edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Puntis, Stephen, Devon Perfect, Abirami Kirubarajan, Sorcha Bolton, Fay Davies, Aimee Hayes, Eli Harriss, and Andrew Molodynski. 2018. "A Systematic Review of Co-Responder Models of Police Mental Health 'street' Triage." *BMC Psychiatry* 18(1):1-11.

Putnam, Israel. 1977. "Poverty Thresholds: Their History and Future Development." Pp. 272-83 in *The Measure of Poverty. Technical Paper I: Documentation of Background Information and Rationale for Current Poverty Matrix,* edited by Mollie Orshansky. Washington, DC: U.S. Department of Health, Education, and Welfare.

Pyrczak, Fred. 2005. *Evaluating Research in Academic Journals: A Practical Guide to Realistic Evaluation,* 3rd ed. Glendale, CA: Pyrczak.

Qiu, Linda. 2020. "Trump's Inaccurate Claims On Hydroxychloroquine." *The New York Times,* May 21.

Quart, Alissa and Yaryna Serkez. 2020. "Who Has Enough Cash to Get Through the Coronavirus Crisis." *The New York Times,* April 26:2.

Quealy, Kevin. 2017. "The Cost of Hurricane Harvey: Only One Recent Storm Comes Close." *The New York Times,* September 1.

Radloff, Lenore. 1977. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement* 1:385-401.

Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies.* Berkeley: University of California Press.

Ragin, Charles C. 1994. *Constructing Social Research.* Thousand Oaks, CA: Sage.

Ragin, Charles C. 2000. *Fuzzy-Set Social Science.* Chicago, IL: University of Chicago Press.

Rainie, Lee and John Horrigan. 2005. *A Decade of Adoption: How the Internet Has Woven Itself into American Life.* Pew Internet & American Life Project. Retrieved from the Pew Internet & American Life Project (www.pewinternet.org/PPF/r/148/report_display.asp [PDF version]).

Rainie, Lee, Aaron Smith, and Maeve Duggan. 2013. *Coming and Going on Facebook.* Washington, DC: Pew Internet & American Life Project.

Rainie, Lee, Joanna Brenner, and Kristen Purcell. 2012. *Photos and Videos as Social Currency Online.* Washington, DC: Pew Internet & American Life Project (http://pewinternet.org/Reports/2012/Online-Pictures.aspx).

Randall, Ann. 2012. "Beneficial Interview Effects in Virtual Worlds: A Case Study." Pp. 131-49 in *Cases in Online Interview Research,* edited by Janet Salmons. Thousand Oaks, CA: Sage.

Rank, Steven G. and Cardell K. Jacobson. 1977. "Hospital Nurses' Compliance with Medication Overdose Orders: A Failure to Replicate." *Journal of Health and Social Behavior* 18:188-93.

Rashbaum, William K. 2002. "Reasons for Crime Drop in New York Elude Many." *The New York Times,* November 29, A28.

Raudenbush, Stephen W. and Robert J. Sampson. 1999. "Ecometrics: Toward a Science of Assessing Ecological Settings, With Application to the Systematic Social Observation of Neighborhoods." *Sociological Methodology* 29:1-41.

Ravitch, Sharon M. and Matthew Riggan. 2016. *Reason & Rigor: How Conceptual Frameworks Guide Research,* 2nd ed. Thousand Oaks, CA: Sage.

Reckdahl, Katy, J. David Goodman, and Chelsea Brasted. 2021. "Levees are Tested as Hurricane Ida Slams Gulf Coast." *The New York Times,* August 30, A1.

Regan, Michael D. 2016. "What Does Voter Turnout Tell Us About the 2016 Election?" *PBS Newshour,* November 20 (http://www.pbs.org/newshour/updates/voter-turnout-2016-elections/).

Reips, Ulf-Dietrich and Christoph Neuhaus. 2002. "WEXTOR: A Web-Based Tool for Generating and Visualising Experimental Designs and Procedures." *Behavior Research Methods Instruments & Computers* 36(2):304-11.

Reisman, David. 1950/1969. *The Lonely Crowd: A Study of the Changing American Character.* New Haven, CT: Yale University Press.

Reiss, Albert J. Jr. 1971a. *The Police and the Public.* New Haven, CT: Yale University Press.

Reiss, Albert J. Jr. 1971b. "Systematic Observations of Natural Social Phenomena." Pp. 3-33 in *Sociological Methodology,* vol. 3, edited by Herbert Costner. San Francisco, CA: Jossey-Bass.

Rele, J. R. 1993. "Demographic Rates: Birth, Death, Marital, and Migration." Pp. 2-1-2-26 in *Readings in Population Research Methodology. Vol. 1, Basic Tools,* edited by Donald J. Bogue, Eduardo E. Arriaga, and Douglas L. Anderton. Chicago: Social Development Center, for the United National Population Fund.

Reuter, Peter, Rosalie Liccardo Pacula, and Jonathan P. Caulkins. 2010. "Addiction Research Centres and the Nurturing of Creativity: RAND's Drug Policy Research Center." *Addiction* 106:253-59.

Rew, Lynn, Deborah Koniak-Griffin, Mary Ann Lewis, Margaret Miles, and Ann O'Sullivan. 2000. "Secondary Data Analysis: New Perspective for Adolescent Research." *Nursing Outlook* 48:223-29.

Reynolds, Paul Davidson. 1979. *Ethical Dilemmas and Social Science Research.* San Francisco, CA: Jossey-Bass.

Richards, J. D. 2022. "Coding, Categorizing, and Theming the Data: A Reflexive Search for Meaning." Pp. 149–167 in *Analyzing and Interpreting Qualitative Research: After the Interview,* edited by C. Vanover, P. Mihas, and J. Saldana. Thousand Oaks, CA: Sage.

Richards, Thomas J. and Lyn Richards. 1994. "Using Computers in Qualitative Research." Pp. 445-62 in *Handbook of Qualitative Research,* edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Richardson, Laurel. 1995. "Narrative and Sociology." Pp. 198-221 in *Representation in Ethnography,* edited by John Van Maanen. Thousand Oaks, CA: Sage.

Riedel, Marc. 2000. *Research Strategies for Secondary Data: A Perspective for Criminology and Criminal Justice.* Thousand Oaks, CA: Sage.

Riessman, Catherine Kohler. 2008. *Narrative Methods for the Human Sciences.* Thousand Oaks, CA: Sage.

Rinehart, Jenny K. and Elizabeth A. Yeater. 2011. "A Qualitative Analysis of Sexual Victimization Narratives." *Violence Against Women* 17(7):925-43.

Ringwalt, Christopher L., Jody M. Greene, Susan T. Ennett, Ronaldo Iachan, Richard R. Clayton, and Carl G. Leukefeld. 1994. *Past and Future Directions of the D.A.R.E. Program: An Evaluation Review.* Research Triangle, NC: Research Triangle Institute.

Rives, Norfleet W. Jr. and William J. Serow. 1988. *Introduction to Applied Demography: Data Sources and Estimation Techniques.* SAGE University Paper Series on Quantitative Applications in the Social Sciences, series No. 07–039. Thousand Oaks, CA: Sage.

Robertson, David Brian. 1993. "The Return to History and the New Institutionalism in American Political Science." *Social Science History* 17:1-36.

Robinson, Eric, Andrew Jones, India Lesser, and Michael Daly. 2021. "International Estimates of Intended Uptake and Refusal of COVID-19 Vaccines: A Rapid Systematic Review and Meta-Analysis of Large Nationally Representative Samples." *Vaccine* 39:2034-2034.

Robinson, Karen A. and Steven N. Goodman. 2011. "A Systematic Examination of the Citation of Prior Research in Reports of Randomized, Controlled Trials." *Annals of Internal Medicine* 154:50-55.

Robinson, Sheila B. and Kimberly Firth Leonard. 2019. *Designing Quality Survey Questions.* New York, NY: Sage.

Rodríguez, Havidán, Joseph Trainor, and Enrico L. Quarantelli. 2006. "Rising to the Challenges of a Catastrophe: The Emergent and Prosocial Behavior Following Hurricane Katrina." *The Annals of the American Academy of Political and Social Science* 604:82-101.

Rogers, Richard. 2019. *Doing Digital Methods.* Thousand Oaks, CA: Sage.

Rogers, Richard. 2021. *Issue Mapping Contextual Essay.* Amsterdam: University of Amsterdam. Accessed December 22, 2021 (http://www.govcom.org/publications/full_list/rogers_issuecrawler_context.pdf).

Roman, Anthony. 2005. *Women's Health Network Client Survey: Field Report.* Unpublished report. Boston: Center for Survey Research, University of Massachusetts.

Romero, Francine Sanders and David W. Romero. 2021. "National Presidential Election Turnout: 1952 to 2020." *American Politics Research* 49 (online) (doi:10.1177/1532673X211031).

Rookey, Bryan D., Steve Hanway, and Don A. Dill-man. 2008. "Does a Probability-Based Household Panel Benefit from Assignment to Postal Response as an Alternative to Internet-Only?" *Public Opinion Quarterly* 72:962-84.

Roper Center for Public Opinion Research. n.d. Retrieved November 18, 2021 (www.ropercenter.cornell.edu).

Rosen, Lawrence. 1995. "The Creation of the Uniform Crime Report: The Role of Social Science." *Social Science History* 19:215-38.

Rosenbach, Margo, Carol Irvin, Angela Merrill, Shanna Shulman, John Czajka, Christopher Trenholm, Susan Williams, So Sasigant Limpa-Amara, and Anna Katz. 2007. *National Evaluation of the State Children's Health Insurance Program: A Decade of Expanding Coverage and Improving Access: Final Report.* Cambridge, MA: Mathematica Policy Research.

Rosenbaum, Dennis P. 2007. "Just Say No to D.A.R.E." *Criminology & Public Policy* 6:815-24.

Rosenbaum, Dennis P. and Gordon S. Hanson. 1998. "Assessing the Effects of School-Based Drug Education: A Six-Year Multi-Level Analysis of Project D.A.R.E." *Journal of Research in Crime and Delinquency* 35:381-412.

Rosenberg, Morris. 1968. *The Logic of Survey Analysis.* New York, NY: Basic Books.

Rosenfeld, Richard. 2004. "The Case of the Unsolved Crime Decline." *Scientific American* 290:82-9.

Rosenthal, Elisabeth. 2000. "Rural Flouting of One-Child Policy Undercuts China's Census." *The New York Times,* April 14, A10.

Rosenthal, Rob. 1994. *Homeless in Paradise: A Map of the Terrain.* Philadelphia, PA: Temple University Press.

Ross, Catherine E. 1990. *Work, Family, and the Sense of Control: Implications for the Psychological Well-Being of Women and Men.* Proposal submitted to the National Science Foundation. Urbana: University of Illinois.

Rossi, Peter H. 1989. *Down and Out in America: The Origins of Homelessness.* Chicago, IL: University of Chicago Press.

Rossi, Peter H. 1999. "Half Truths With Real Consequences: Journalism, Research, and Public Policy. Three Encounters." *Contemporary Sociology* 28:1-5.

Rossi, Peter H. and Howard E. Freeman. 1989. *Evaluation: A Systematic Approach,* 4th ed. Newbury Park, CA: Sage.

Rossman, Gretchen B. and Sharon F. Rallis. 1998. *Learning in the Field: An Introduction to Qualitative Research.* Thousand Oaks, CA: Sage.

Roth, Dee. 1990. "Homelessness in Ohio: A Statewide Epidemiological Study." Pp. 145-63 in *Homeless in the United States, Vol. 1: State Surveys,* edited by Jamshid Momeni. New York, NY: Greenwood.

Rothbaum, Jonathan and Adam Bee. 2021. "Coronavirus Infects Surveys, Too: Survey Nonresponse Bias and the Coronavirus Pandemic *."

Rotolo, Thomas and Charles R. Tittle. 2006. "Population Size, Change, and Crime in U.S. Cities." *Journal of Quantitative Criminology* 22:341-67.

Rubin, Herbert J. and Irene S. Rubin. 1995. *Qualitative Interviewing: The Art of Hearing Data.* Thousand Oaks, CA: Sage.

Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens. 1992. *Capitalist Development and Democracy.* Chicago, IL: University of Chicago Press.

Ruggles, Patricia. 1990. *Drawing the Line: Alternative Poverty Measures and Their Implications for Public Policy.* Washington, DC: Urban Institute Press.

Russonello, Giovanni. 2020. "Do Voters Want to 'Defund the Police'? Depends on What 'Defund' Means." *The New York Times,* 17.

Sacks, Stanley, Karen McKendrick, George DeLeon, Michael T. French, and Kathryn E. McCollister. 2002. "Benefit-Cost Analysis of a Modified Therapeutic Community for Mentally Ill Chemical Abusers." *Evaluation and Program Planning* 25:137-48.

Saksena, Jyotika and Shannon L. McMorrow. 2020. "Through Their Eyes: A Photovoice and Interview Exploration of Integration Experiences of Congolese Refugee Women in Indianapolis." *Journal of International Migration and Integration* 21(2):529-49.

Saldana, Johnny. 2021. *The Coding Manual for Qualitative Researchers.* 4th ed. Thousand Oaks, CA: Sage.

Saldana, Johnny and Matt Omasta. 2021. *Qualitative Research.* 2nd ed. New York, NY: Sage.

Saldana, Johnny and Matt Omasta. 2022. *Qualitative Research: Analyzing Life.* 2nd ed. Thousand Oaks, CA: Sage.

Salisbury, Robert H. 1975. "Research on Political Participation." *American Journal of Political Science* 19(May):323-41.

Salmons, Janet. 2012. "Designing and Conducting Research With Online Interviews." Pp. 1-35 in *Cases in Online Interview Research,* edited by Janet Salmons. Thousand Oaks, CA: Sage.

Salmons, Janet E. 2016. *Doing Qualitative Research Online.* Thousand Oaks, CA: Sage.

Sampson, Robert J. 1987. "Urban Black Violence: The Effect of Male Joblessness and Family Disruption." *American Journal of Sociology* 93(September):348-82.

Sampson, Robert J. and Alix S. Winter. 2018. "Poisoned Development: Assessing Childhood Lead Exposure as a Cause of Crime in a Birth Cohort Followed Through Adolescence." *Criminology* 56:269-301.

Sampson, Robert J. and John H. Laub. 1990. "Crime and Deviance Over the Life Course: The Salience of Adult Social Bonds." *American Sociological Review* 55(October):609-27.

Sampson, Robert J. and John H. Laub. 1993. "Structural Variations in Juvenile Court Processing: Inequality, the Underclass, and Social Control." *Law & Society Review* 27(2):285-311.

Sampson, Robert J. and John H. Laub. 1994. "Urban Poverty and the Family Context of Delinquency: A New Look at Structure and Process in a Classic Study." *Child Development* 65:523-40.

Sampson, Robert J. and Stephen W. Raudenbush. 1999. "Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods." *American Journal of Sociology* 105:603-651.

Sampson, Robert J. and Stephen W. Raudenbush. 2001. "Disorder in Urban Neighborhoods—Does It Lead to Crime?" In *Research in Brief.* Washington, DC: National Institute of Justice, U.S. Department of Justice.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277:918-24.

Sampson, Robert J., William Julius Wilson, and Hanna Katz. 2018. "Reassessing Toward a Theory of Race, Crime, and Urban Inequality." *Du Bois Review* 15(1):13-34.

Sanchez, Gabriel R., Francisco I. Pedraza, and Edward D. Vargas. n.d. "The Impact of the Affordable Care Act on Latino Access to Health Insurance A Publication of the RWJF Center for Health Policy." Albuquerque: Center for Health Policy, University of New Mexico. https://healthpolicy.unm.edu/sites/default/files/ACA_2015_Research_Brief_FINAL.pdf

Sandstrom, Heather, Sarah Benatar, Erica Greenberg, Ian Hill, Embry Howell, Timothy Triplett, Margaret Wilkinson, Todd Franke, Christina Christie, Maria Lourdes Brown, and Jennifer Ho. 2015. *Welcome Baby Home Visiting: Findings From the 36-Month Child & Family Survey and 3-Year Longitudinal Results.* Urban Institute (http://www.first5la.org/postfiles/files/36-Month%20Outcomes%20Report.pdf).

Saris, Willem E., Melanie Revilla, Jon A. Krosnick, and Eric M. Shaeffer, E. 2010. "Comparing Questions with Agree–Disagree Response Options to Questions with Item-Specific Response Options." *Survey Research Methods,* 4(1):61-79.

Sarkar, Sucharita. 2021. "Of Late Alarms, Long Queues, and Online Attendances: My Experiences of COVID Time." *Qualitative Inquiry* 27(7):820-23.

Sayer, Andrew. 2003. "Reductionism in Social Science." Lancaster, England: Department of Sociology, Lancaster University. Retrieved from (www.lancaster.ac.uk/fass/sociology/research/publications/papers/sayer-paris1.pdf).

Schaeffer, Nora Cate and Stanley Presser. 2003. "The Science of Asking Questions." *Annual Review of Sociology* 29:65-88.

Scharf, Darcell P., Katherine J. Mathews, Pamela Jackson, Jonathan Hofsuemmer, Emeobong Martin, and Dorothy Edwards. 2010. "More Than Tuskegee: Understanding Mistrust About Research Participation." *Journal of Health Care for the Poor and Underserved* 21(3):879-97.

Schegloff, Emanuel A. 1996. "Issues of Relevance for Discourse Analysis: Contingency in Action, Interaction and Coparticipant Context." Pp. 3-35 in *Computational and Conversational Discourse: Burning Issues—An Interdisciplinary Account,* edited by Eduard H. Hovy and Donia R. Scott. New York, NY: Springer.

Schleyer, Titus K. L. and Jane L. Forrest. 2000. "Methods for the Design and Administration of Web-Based Surveys." *Journal of the American Medical Informatics Association* 7:418-25.

Schober, Michael F. 1999. "Making Sense of Survey Questions." Pp. 77-94 in *Cognition and Survey Research,* edited by Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwartz, Judith M. Tanur, and Roger Tourangeau. New York, NY: Wiley.

Schofield, Janet Ward. 2002. "Increasing the Generalizability of Qualitative Research." Pp. 171-203 in *The Qualitative Researcher's Companion,* edited by A. Michael Huberman and Matthew B. Miles. Thousand Oaks, CA: Sage.

Schorr, Lisbeth B. and Daniel Yankelovich. 2000. "In Search of a Gold Standard for Social Programs." *The Boston Globe,* February 18, A19.

Schreck, Christopher J., Eric A. Steward, and Bonnie S. Fisher. 2006. "Self-Control, Victimization, and Their Influence on Risky Lifestyles: A Longitudinal Analysis Using Panel Data." *Journal of Quantitative Criminology* 22:319-40.

Schreer, George E., Saundra Smith, and Kirsten Thomas. 2009. "'Shopping While Black': Examining Racial Discrimination in a Retail Setting." *Journal of Applied Social Psychology* 39:1432-44.

Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.* New York, NY: Academic Press.

Schutt, Russell K. 1986. *Organization in a Changing Environment.* Albany, NY: State University of New York Press.

Schutt, Russell K. 2011a. "Evaluation of the Coordinated Care Program." Proposal to Institutional Review Board for the Protection of Human Subjects, University of Massachusetts Boston. Boston: University of Massachusetts Boston.

Schutt, Russell K. 2011b. *Homelessness, Housing, and Mental Illness.* Cambridge, MA: Harvard University Press.

Schutt, Russell K. (with the assistance of Tatjana Meschede). 1992. *The Perspectives of DMH Shelter Staff: Their Clients, Their Jobs, Their Shelters and the Service System.* Unpublished report to the Metro Boston Region of the Massachusetts Department of Mental Health. Boston, MA: Department of Sociology, University of Massachusetts.

Schutt, Russell K. and Jacqueline Fawcett. 2005. *Case Management in the Women's Health Network.* Boston: University of Massachusetts. Unpublished report.

Schutt, Russell K. and M. L. Fennell. 1992. "Shelter Staff Satisfaction with Services, the Service Network and Their Jobs." *Current Research on Occupations and Professions* 7:177-200.

Schutt, Russell K. and Stephen M. Goldfinger. 1996. "Housing Preferences and Perceptions of Health and Functioning Among Homeless Mentally Ill Persons." *Psychiatric Services* 47:381-86.

Schutt, Russell K. and W. Dale Dannefer. 1988. "Detention Decisions in Juvenile Cases: JINS, JDs and Gender." *Law & Society Review* 22(3):509-20.

Schutt, Russell K., Daniel T. O'Brien, Lee Hargraves, Dan Dooley, Floyd Fowler, Ryan Qi Want, Alina Ristea, Anthony Roman, Mehrnaz Amiri, Sage Gibbons, Hannah Grabowski, and Nikola Kovacevic. 2020. *Living in Boston During COVID-19: Economic Strains.* Boston: Boston Area Research Initiative (BARI), Northeastern University; BEACON Project, Center for Survey Research, University of Massachusetts Boston; Boston Public Health Commission, City of Boston.

Schutt, Russell K., Richard L. Hough, Stephen M. Goldfinger, Anthony F. Lehman, David L. Shern, Elie S. Valencia, and Patricia A. Wood. 2009. "Lessening Homelessness Among Persons With Mental Illness: A Comparison of Five Randomized Treatment Trials." *Asian Journal of Psychiatry* 2:100-05 (www.ncbi.nlm.nih.gov/pmc/articles/PMC2788308/pdf/nihms147743.pdf).

Schutt, Russell K., Stephen M. Goldfinger, and Walter E. Penk. 1997. "Satisfaction With Residence and With Life: When Homeless Mentally Ill Persons Are Housed." *Evaluation and Program Planning* 20(2):185-94.

Schutt, Russell K., Suzanne Gunston, and John O'Brien. 1992. "The Impact of AIDS Prevention Efforts on AIDS Knowledge and Behavior Among Sheltered Homeless Adults." *Sociological Practice Review* 3(1):1-7.

Schutt, Russell K., Xiaogang Deng, Gerald R. Garrett, Stephanie Hartwell, Sylvia Mignon, Joseph Bebo, Matthew O'Neill, Mary Aruda, Pat Duynstee, Pam DiNapoli, and Helen Reiskin. 1996. *Substance Use and Abuse Among UMass Boston Students.* Boston, MA: Department of Sociology, University of Massachusetts. Unpublished report.

Schwandt, Thomas A. 1994. "Constructivist, Interpretivist Approaches to Human Inquiry." Pp. 118-37 in *Handbook of Qualitative Research,* edited by Norman K. Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: SAGE.

Schwartz, John. 2005. "Myths Run Wild in Blog Tsunami Debate." *The New York Times,* January 3, A9.

Schwarz, Norbert. 2010. "Measurement as Cooperative Communication: What Research Participants Learn From Questionnaires." Pp. 43-59 in *The SAGE Handbook of Measurement,* edited by Geoffrey Walford, Eric Tucker, and Madhu Viswanathan. Thousand Oaks, CA: Sage.

Scott, Janny. 2001. "A Nation by the Numbers, Smudged." *The New York Times,* July 1, 21, 22.

Scott, John. 2013. *Social Network Analysis,* 3rd ed. Thousand Oaks, CA: Sage.

Scriven, Michael. 1972a. "The Methodology of Evaluation." Pp. 123-36 in *Evaluating Action Programs: Readings in Social Action and Education,* edited by Carol H. Weiss. Boston: Allyn & Bacon.

Scriven, Michael. 1972b. "Prose and Cons About Goal-Free Evaluation." *Evaluation Comment* 3:1-7.

Scull, Andrew T. 1988. "Deviance and Social Control." Pp. 667-93 in *Handbook of Sociology,* edited by Neil J. Smelser. Newbury Park, CA: Sage.

Sechrest, Lee and Souraya Sidani. 1995. "Quantitative and Qualitative Methods: Is There an Alternative?" *Evaluation and Program Planning* 18:77-87.

Sechrist, Stacy M. and John D. Weil. 2017. "Assessing the Impact of a Focused Deterrence Strategy to Combat Intimate Partner Domestic Violence." *Violence Against Women.* Online first.

Seddighi, Hamed, Ibrahim Salmani, Mohammad Hossein Javadi, and Saeideh Seddighi. 2019. "Child Abuse in Natural Disasters and Conflicts: A Systematic Review." *Trauma, Violence & Abuse* 22(1):176-75 (https://doi.org/10.1177/1524838019835973).

Selm, Martine Van and Nicholas W. Jankowski. 2006. "Conducting Online Surveys." *Quality & Quantity* 40:435-56.

Selwitz, Ada Sue, Norma Epley, and Janelle Erickson. 2013. *Basic Institutional Review Board (IRB): Regulations and Review Process.* Miami, FL: Collaborative Institutional Training Initiative at the University of Miami (https://www.citiprogram.org/).

Shadish, William R., Thomas D. Cook, and Laura C. Leviton (Eds.). 1991. *Foundations of Program Evaluation: Theories of Practice.* Thousand Oaks, CA: Sage.

Sharkey, Patrick, Gerard Torrats-Espinosa, and Delaram Takyar. 2017. "Community and the Crime Decline: The Causal Effect of Local Nonprofits on Violent Crime." *American Sociological Review* 82(6):1214-40.

Sharma, Divya. 2009. "Research Ethics and Sensitive Behaviors: Underground Economy." Pp. 426-41 in *The Handbook of Social Research Ethics,* edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Shepherd, Jane, David Hill, Joel Bristor, and Pat Montalvan. 1996. "Converting an Ongoing Health Study to CAPI: Findings From the National Health and Nutrition Study." Pp. 159–164 in *Health Survey Research Methods Conference Proceedings,* edited by Richard B. Warnecke. Hyattsville, MD: U.S. Department of Health and Human Services.

Sherman, Lawrence W. 1992. *Policing Domestic Violence: Experiments and Dilemmas.* New York, NY: Free Press.

Sherman, Lawrence W. 1993. "Implications of a Failure to Read the Literature." *American Sociological Review* 58:888-89.

Sherman, Lawrence W. 2020. "Evidence-Based Policing and Fatal Police Shootings: Promise, Problems, and Prospects." *Annals of the American Academy of Political and Social Science* 687(1):8-26.

Sherman, Lawrence W. and Douglas A. Smith, with Janell D. Schmidt and Dennis P. Rogan. 1992. "Crime, Punishment, and Stake in Conformity: Legal and Informal Control of Domestic Violence." *American Sociological Review* 57:680-90.

Sherman, Lawrence W. and Heather M. Harris. 2013. "Increased Homicide Victimization of Suspects Arrested for Domestic Assault: A 23-Year Follow-Up of the Milwaukee Domestic Violence Experiment (MilDVE). *Journal of Experimental Criminology* 9:491-514.

Sherman, Lawrence W. and Richard A. Berk. 1984. "The Specific Deterrent Effects of Arrest for Domestic Assault." *American Sociological Review* 49:261-72.

Shermer, Michael. 1997. *Why People Believe Weird Things: Pseudoscience, Superstition, and Other Confusions of Our Time.* New York, NY: W. H. Freeman.

Shrout, Patrick E. 2011. "Integrating Causal Analysis into Psychopathology Research." Pp. 3-24 in *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures,* edited by Patrick E. Shrout, Katherine M. Keyes, and Katherine Ornstein. New York, NY: Oxford University Press.

Shrum, Wesley and Greg Scott. 2017. *Video Ethnography in Practice: Planning, Shooting, and Editing for Social Analysis.* Thousand Oaks, CA: Sage.

Sieber, Joan E. 1992. *Planning Ethically Responsible Research: A Guide for Students and Internal Review Boards.* Thousand Oaks, CA: Sage.

Sieber, Joan E. and Martin B. Tolich. 2013. *Planning Ethically Responsible Research,* 2nd ed. Thousand Oaks, CA: Sage.

Silva, Fabiana. 2018. "The Strength of Whites' Ties: How Employers Reward the Referrals of Black and White Jobseekers." *Social Forces* 97(2):741-68.

Sin, Chih Hoong. 2005. "Seeking Informed Consent: Reflections on Research Practice." *Sociology* 39:277-94.

Sjoberg, Gideon (Ed.). 1967. *Ethics, Politics, and Social Research.* Cambridge, MA: Schenkman.

Skinner, Harvey A. and Wen-Jenn Sheu. 1982. "Reliability of Alcohol Use Indices: The Lifetime Drinking History and the MAST." *Journal of Studies on Alcohol* 43(11):1157-70.

Skocpol, Theda and Margaret Somers. 1979. "The Uses of Comparative History in Macrosocial Inquiry." Pp. 72-95 in *Social Revolutions in the Modern World,* edited by Theda Skocpol. New York, NY: Cambridge University Press.

Skocpol, Theda. 1979. States and Social Revolutions: A Comparative Analysis of France, Russia, and China. New York: Cambridge University Press.

Skocpol, Theda. 1984. "Emerging Agendas and Recurrent Strategies in Historical Sociology." Pp. 356-91 in *Vision and Method in Historical Sociology,* edited by Theda Skocpol. New York, NY: Cambridge University Press.

Slater, Mel, Angus Antley, Adam Davison, David Swapp, Christoph Guger, Chris Barker, Nancy Pistrang, and Maria V. Sanchez-Vives. 2006. "A Virtual Reprise of the Stanley Mil-gram Obedience Experiments." *PLoS ONE,* 1(1).

Sloboda, Zili, Richard C. Stephens, Peggy C. Stephens, Scott F. Grey, Brent Teasdale, Richard D. Hawthorne, Joseph Williams, and Jesse F. Marquette. 2009. "The Adolescent Substance Abuse Prevention Study: A Randomized Field Trial of a Universal Substance Abuse Prevention Program." *Drug and Alcohol Dependence* 102(1–3):1-10.

Smith, Erica L. and Alexia Cooper. 2013. *Homicide in the U.S. Known to Law Enforcement, 2011.* Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice.

Smith, Tom W. 1984. "Nonattitudes: A Review and Evaluation." Pp. 215–255 in *Surveying Subjective Phenomena,* vol. 2, edited by Charles F. Turner and Elizabeth Martin. New York: Russell Sage Foundation.

Smith, Tom W. 1995. "Trends in Nonresponse Rates." *International Journal of Public Opinion Research* 7: 157-71.

Smith, Tom W., Michael Davern, Jeremy Freese, and Michael Hout. 2017. *General Social Surveys, 1972–2016.* [machine-readable data file]. Principal Investigator, Tom W. Smith; Co-Principal Investigators, Peter V. Marsden and Michael Hout, NORC ed. Chicago: NORC. 1 data file (62,466 logical records) and 1 codebook (3,689 pp).

Smithson, Janet. 2008. "Focus Groups." Pp. 357-70 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Smyth, Jolene D., Don A. Dillman, Leah Melani Christian, and Michael J. Stern. 2004. *How Visual Grouping Influences Answers to Internet Surveys.* Extended version of paper presented at the Annual Meeting of the American Association for Public Opinion Research, Phoenix, AZ, May 13 (http://survey.sesrc.wsu.edu/dillman/papers.htm).

Snipp, C. Matthew. 2003. "Racial Measurement in the American Census: Past Practices and Implications for the Future." *Annual Review of Sociology* 29:563-88.

Snow, David L., Jacob Kraemer Tebes, and Michael W. Arthur. 1992. "Panel Attrition and External Validity in Adolescent Substance Use Research." *Journal of Consulting and Clinical Psychology* 60:804-07.

Sobell, Linda C., Mark B. Sobell, Diane M. Riley, Reinhard Schuller, D. Sigfrido Pavan, Anthony Cancilla, Felix Klajner, and Gloria I. Leo. 1988. "The Reliability of Alcohol Abusers' Self-Reports of Drinking and Life Events That Occurred in the Distant Past." *Journal of Studies on Alcohol* 49(2):225-32.

Social Trends Institute. 2017. *The Sustainable Demographic Dividend: What Do Marriage & Fertility Have to Do With the Economy?* (http://sustaindemographicdividend.org/articles/international-family-indicators/global-family-structure).

Societal Experts Action Network COVID-19 Survey Archive. n.d. "Welcome to the Societal Experts Action Network (SEAN) COVID-19 Survey Archive." Retrieved November 18, 2021 (https://covid-19.parc.us.com/client/index.html#/).

Sorowski, Piotr, Emanuel Kulczycki, Agnieszka Sorokowska, and Katarzyna Pisanski. 2017. "Predatory Journals Recruit Fake Editor." *Nature* 543 (March 23):481-83.

Sosin, Michael R., Paul Colson, and Susan Grossman. 1988. *Homelessness in Chicago: Poverty and Pathology, Social Institutions and Social Change.* Chicago, IL: Chicago Community Trust.

Sousa, William H., James R. Coldren, Denise Rodriguez, and Anthony A. Braga. 2016. "Research on Body Worn Cameras: Meeting the Challenges of Police Operations, Program Implementation, and Randomized Controlled Trial Designs." *Police Quarterly* 19(3):363-84.

Southall, Ashley. 2020. "Why a Drop in Domestic Violence Reports." *The New York Times,* April 19, A17.

Speiglman, Richard and Patricia Spear. 2009. "The Role of Institutional Review Boards: Ethics: Now You See Them, Now You Don't." Pp. 121-34 in *The Handbook of Social Research Ethics,* edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Spretnak, Charlene. 1991. *States of Grace: The Recovery of Meaning in the Postmodern Age.* New York, NY: HarperCollins.

St. Jean, Peter K. B. 2007. *Pockets of Crime: Broken Windows, Collective Efficacy, and the Criminal Point of View.* Chicago, IL: University of Chicago Press.

St. Pierre, Robert G. and Peter H. Rossi. 2006. "Randomize Groups, Not Individuals: A Strategy for Improving Early Childhood Programs." *Evaluation Review* 30:656-85.

Stake, Robert and Fazal Rizvi. 2009. "Research Ethics in Transnational Spaces." Pp. 521-36 in *The Handbook of Social Research Ethics,* edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Stake, Robert E. 1995. *The Art of Case Study Research.* Thousand Oaks, CA: Sage.

Stephen, Jason M., Michael F. Young, and Thomas Calabrese. 2007. "Does Moral Judgment Go Offline When Students Are Online? A Comparative Analysis of Undergraduates' Beliefs and Behaviors Related to Conventional and Digital Cheating." *Ethics & Behavior* 17:233-54.

Stewart, David W. and Michael A. Kamins. 1993. *Secondary Research: Information Sources and Methods,* 2nd ed. Newbury Park, CA: Sage.

Stewart, Nikita. 2017. "As More Opt for Streets, City Offers a Place to Go." *The New York Times,* July 19, A20.

Stewart, Nikita. 2020. "Child Abuse Cases Drop 51 Percent. The Authorities Are Very Worried." *The New York Times,* June 9 (updated Aug. 7).

Stille, Alexander. 2000. "A Happiness Index with a Long Reach: Beyond G.N.P. to Subtler Measures." *The New York Times,* May 20, A17, A19.

Stokoe, Elizabeth. 2006. "On Ethnomethodology, Feminism, and the Analysis of Categorical Reference to Gender in Talk-in-Interaction." *The Sociological Review* 54:467-94.

Strahan, Robert and Kathleen Carrese Gerbasi. 1972. "Short, Homogeneous Versions of the Marlowe-Crowne Social Desirability Scale." *Journal of Clinical Psychology* 28:191-93.

Strauss, Anselm L. and Juliette Corbin. 1990. *The Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* Newbury Park, CA: Sage.

Strunk, William Jr. and E. B. White. 2000. *The Elements of Style,* 4th ed. New York, NY: Allyn & Bacon.

Sudman, Seymour. 1976. *Applied Sampling.* New York, NY: Academic Press.

Sue, Valerie M. and Lois A. Ritter. 2012. *Conducting Online Surveys,* 2nd ed. Thousand Oaks, CA: Sage.

Sugie, Naomi F., Noah D. Zatz, and Dallas Augustine. 2020. "Employer Aversion to Criminal Records: An Experimental Study of Mechanisms." *Criminology* 58(1):5-34.

Sulkunen, Pekka. 2008. "Social Research and Social Practice in Post-Positivist Society." Pp. 68-80 in *The SAGE Handbook of Social Research Methods,* edited by Pertti Alasuutari, Leonard Bickman, and Julia Brannen. Thousand Oaks, CA: Sage.

Sunderland, Antonia. 2005. *Children, Families and Welfare Reform: A Three-City Study.* Princeton, NJ: The Robert Wood Johnson Foundation.

"Survey on Adultery: 'I Do' Means 'I Don't.'" 1993. *The New York Times,* October 19, A20.

Survey Research Laboratory. 2008. *List of Academic and Not-for-Profit Survey Research Organizations (LANSRO).* Chicago, IL: College of Urban Planning and Public Affairs, University of Illinois at Chicago (www.srl.uic.edu/LANSRO.doc).

SurveyMonkey. n.d. "Engaged Survey Respondents, Honest Feedback" (https://www.surveymonkey.com/mp/audience/our-survey-respondents).

Swarns, Rachel L. 2004. "Hispanics Debate Racial Grouping by Census." *The New York Times,* October 24, A1, A18.

Taub, Amanda. 2020. "A New Covid-19 Crisis: Domestic Abuse Rises Worldwide." The New York Times, April 6 (update April 14) (https://www.nytimes.com/2020/06/09/world/coronavirus-domestic-violence.html).

Tavernise, Sabrina. 2002. "How Many Russians? Let Us Weigh the Count, Cooperation or No." *The New York Times,* October 10, A13.

Tavernise, Sabrina and Richard A. Oppel Jr. 2020. "Spit On, Yelled At, Attacked: Chinese-Americans Fear for Their Safety." *The New York Times,* March 23.

Tavory, Iddo and Stefan Timmermans. 2014. *Abductive Analysis: Theorizing Qualitative Research.* Chicago, IL: University of Chicago Press.

Taylor, Charles Lewis and David A. Jodice. 1986. *World Handbook of Political and Social Indicators III: 1948–1982.* File available from the Inter-university Consortium for Political and Social Research (ICPSR), Study #7761.

Taylor, Jerry. 1999. "DARE Gets Updated in Some Area Schools, Others Drop Program." *The Boston Sunday Globe,* May 16, 1, 11.

Testa, Maria, Jennifer A. Livingston, and Carol VanZile-Tamsen. 2011. "Advancing the Study of Violence Against Women Using Mixed Methods: Integrating Qualitative Methods Into a Quantitative Research Program." *Violence Against Women* 17(2):236-50.

Thomas, Neil. 2005. "Disaster Center Researchers to Study Katrina Response." *UDaily,* September 15 (www.udel.edu/PR/UDaily/2005/mar/DRC091505.html).

Thompson, Clive. 2019. "The Secret History of Women in Coding." *The New York Times Magazine,* February 17:38-45, 54-55.

Thorne, Barrie. 1993. *Gender Play: Girls and Boys in School.* New Brunswick, NJ: Rutgers University Press.

Thorsen, Maggie L. and Jennifer Pearce-Morris. 2016. "Adolescent Mental Health and Dating in Young Adulthood." *Society and Mental Health* 6(3):223-45.

Thrasher, Frederic M. 1927. *The Gang: A Study of 1,313 Gangs in Chicago.* Chicago, IL: University of Chicago Press.

Tierolf, Bas, Edith Geurts, and Majone Steketee. 2021. "Domestic Violence in Families in the Netherlands During the Coronavirus Crisis: A Mixed Method Study." *Child Abuse & Neglect* 116:10480 (https://doi.org/10.1016/j.chiabu.2020.104800).

Tigges, Beth Baldwin. 2003. "Parental Consent and Adolescent Risk Behavior Research." *Journal of Nursing Scholarship* 35:283-89.

Tight, Malcolm. 2017. *Understanding Case Study Research: Small-Scale Research with Meaning.* Thousand Oaks, CA: Sage.

Timmer, Doug A., D. Stanley Eitzen, and Kathryn D. Talley. 1993. *Paths to Homelessness: Extreme Poverty and the Urban Housing Crisis.* Boulder, CO: Westview Press.

Tinkler, Penny. 2013. *Using Photographs in Social and Historical Research.* Thousand Oaks, CA: Sage.

Tjaden, Patricia and Nancy Thoennes. 2000. *Extent, Nature, and Consequences of Intimate Partner Violence: Findings from the National Violence Against Women Survey, NCJ 181867.* Washington, DC: Office of Justice Programs, National Institute of Justice and the Centers for Disease Control and Prevention.

Toby, Jackson. 1957. "Social Disorganization and Stake in Conformity: Complementary Factors in the Predatory Behavior of Hoodlums." *Journal of Criminal Law, Criminology and Police Science* 48:12-7.

Toepoel, Vera, Alerk Amin, and Peter Lugtig. n.d. "The Effects of Adding a Mobile-Compatible Design to the American Life Panel." RAND Labor and Population, American Life Panel (http://slideplayer.com/slide/9015836/).

Toepoel, Vera. 2016. *Doing Surveys Online.* Thousand Oaks, CA: Sage.

Toppo, Greg. 2002. "Antidrug Program Backed by Study." *The Boston Globe,* October 29, A10.

Tourangeau, Roger, Frederick G. Conrad, and Mick P. Couper. 2012. *The Science of Web Surveys.* Oxford: Oxford University Press.

Tourangeau, Roger. 1999. "Context Effects." Pp. 111-32 in *Cognition and Survey Research,* edited by Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwartz, Judith M. Tanur, and Roger Tourangeau. New York, NY: Wiley.

Tourangeau, Roger. 2004. "Survey Research and Societal Change." *Annual Review of Psychology* 55:775-801.

Townsend, Meg, Dana Hunt, Caity Baxter, and Peter Finn. 2005. *Interim Report: Evaluability Assessment of the President's Family Justice Center Initiative.* Cambridge, MA: Abt Associates (www.ncjrs.gov/pdffiles1/nij/grants/212278.pdf)

Tufte, Edward R. 1983. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

Turabian, Kate L. 2007. *A Manual for Writers of Term Papers, Theses, and Dissertations,* 7th ed. Chicago, IL: University of Chicago Press.

Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less From Each Other.* New York, NY: Basic Books.

Turkle, Sherry. 2015. *Reclaiming Conversation: The Power of Talk in a Digital Age.* New York, NY: Penguin Books.

Turner, Charles F. and Elizabeth Martin (Eds.). 1984. *Surveying Subjective Phenomena,* vols. 1 and 2. New York, NY: Russell Sage Foundation.

Turner, Jonathan H., Leonard Beeghley, and Charles H. Powers. 1995. *The Emergence of Sociological Theory,* 3rd ed. Belmont, CA: Wadsworth.

Tyler, Tom R. 1990. "The Social Psychology of Authority: Why Do People Obey an Order to Harm Others?" *Law & Society Review* 24:1089-102.

Tyson, Alec. 2021. "Republicans Remain Far Less Likely Than Democrats to View COVID-19 as a Major Threat to Public Health." *Pew Research Center,* July 22.

U.S. Bureau of Economic Analysis. 2004. *Customer Satisfaction Survey Report, FY 2004.* Washington, DC: U.S Department of Commerce.

U.S. Bureau of Labor Statistics, Department of Labor. 1991. *Major Programs of the Bureau of Labor Statistics.* Washington, DC: U.S. Bureau of Labor Statistics, Department of Labor.

U.S. Bureau of Labor Statistics, Department of Labor. 1997a. *Employment and Earnings.* Washington, DC: U.S. Bureau of Labor Statistics, Department of Labor.

U.S. Bureau of Labor Statistics, Department of Labor. 1997b. *Handbook of Methods.* Washington, DC: U.S. Bureau of Labor Statistics, Department of Labor.

U.S. Bureau of Labor Statistics, Department of Labor. 2014. *National Longitudinal Survey of Youth: Children and Young Adults* (https ://www.nlsinfo.org/content/cohorts/nlsy79-children/other-documentation/questionnaires).

U.S. Bureau of Labor Statistics. 2022. *NLSY79 National Longitudinal Survey of Youth, Child and Young Adult Codebook Supplement*: Author. https://www.nlsinfo.org/content/cohorts/nlsy79-children/topical-guide/attitudes/attitudes-expectations

U.S. Census Bureau. 1981. *Section 1, Vital Statistics. Statistical Abstract of the United States, 1981,* 102nd ed. Washington, DC: U.S. Department of Commerce, Census Bureau.

U.S. Census Bureau. 1994. *Census Catalog and Guide, 1994.* Washington, DC: U.S. Bureau of the Census.

U.S. Census Bureau. 1999. *United States Census 2000, Updated Summary: Census 2000 Operational Plan.* Washington, DC: U.S. Department of Commerce, Bureau of the Census, February.

U.S. Census Bureau. 2001. "Statement by William G. Barron Jr. on the Current Status of Results of Census 2000 Accuracy and Coverage Evaluation Survey." *United States Department of Commerce News,* July 13 (www.census.gov/PressRelease/www/2001 /cb01cs06.html).

U.S. Census Bureau. 2006. *Census Bureau Guideline: Language Translation of Data Collection Instruments and Supporting Materials.* Washington, DC: U.S. Census Bureau, Census Advisory Committees (www.census.gov/cac/www/007585.html).

U.S. Census Bureau. 2010a. "$1.6 Billion in 2010 Census Savings Returned." *United States Department of Commerce News,* August 10 (http://2010.census.gov/news/releases/operations/).

U.S. Census Bureau. 2010b. "Door-to-Door Visits Begin for 2010 Census." *United States Department of Commerce News,* April 30 (http://2010.census.gov/news/releases/operations/).

U.S. Census Bureau. 2010c. "2010 Census Forms Arrive in 120 Million Mailboxes Across Nation." *United States Department of Commerce News,* March 15 (http://2010.census.gov/news/releases/operations/).

U.S. Census Bureau. 2010d. *Questions and Answers: Real People, Real Questions, Real Answers.* Washington, DC: U.S. Department of Commerce, Bureau of the Census (http://2010.census.gov/2010census/about/answers.php).

U.S. Census Bureau. 2013. "Census Bureau News— Extended Measures of Well-Being: Living Conditions in the United States: 2011." Washington, DC: U.S. Census Bureau (www.prnewswire.com/news-releases/census-bureau-news----extended-measu resof-well-being-living-conditions-in-theunited-states-2011-222530671.html).

U.S. Census Bureau. 2014. "Foreign-Born: 2014 Current Population Survey Detailed Tables" (https://www.census.gov/data/tables/2014/demo/foreign-born/cps2014.html).

U.S. Census Bureau. 2020a. "2020 Census: Tracking Self-Response and Nonresponse Follow up for Housing Units by State" (https://www.census.gov/library/visualizations/interactive/2020-census-total-response-rates-by-state.html).

U, S. Census Bureau. 2020b. *Household Pulse Survey* (https://www2.census.gov/programs-surveys/demo/technical-documentation/hhp/Phase_3-2_Household_Pulse_Survey_FINAL_English_SKIPS_081821.pdf).

U.S. Census Bureau. 2020c. *Selected Social Characteristics in the United States, Table DP02.* Retrieved November 18, 2021 (https://data.census.gov/cedsci/table?q=foreign%20born&tid=ACSDP1Y2019.DP02).

U. S. Census Bureau. 2020d. "Voting and Registration in the Election of November 2020." Retrieved November 18, 2021 (https://www.census.gov/data/tables/time-series/demo/voting-and-registration/p20-585.html).

U.S. Census Bureau. 2021. "Measuring Household Experiences during the Coronavirus Pandemic" (https://www.census.gov/data/experimental-data-products/household-pulse-survey.html).

U.S. Census Bureau. 2021. *Race and Ethnicity in the United States: 2010 Census and 2020 Census.* Washington DC: U.S. Department of Commerce, Census Bureau.

U.S. Census Bureau. n.d. "About the 2020 Census" (https://www.census.gov/programs-surveys/decennial-census/decade/2020/about.html).

U.S. Census Bureau. n.d.a. "American Community Survey (ACS)" (https://www.census.gov/programs-surveys/acs.html)

U.S. Census Bureau. n.d.b. "Philadelphia City, Philadelphia County, Pennsylvania" (https://data.census.gov/cedsci/profile?g=0600000US4210160000).

U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services. 1995. *Client-Level Evaluation Procedure Manual.* Washington, DC: U.S. Department of Health and Human Services.

U.S. National Library of Medicine. n.d. "ClinicalTrials.gov: History, Policies, and Law" (https://clinicaltrials.gov/ct2/about-site/history).

U.S. Office of Management and Budget. 2002. *Government and Performance Results Act of 1993.* Washington, DC: U.S. Office of Management and Budget, Executive Office of the President.

Uchitelle, Louis. 1997. "Measuring Inflation: Can't Do It, Can't Stop Trying." *The New York Times,* March 16, 4.

Uchitelle, Louis. 1999. "Devising New Math to Define Poverty." *The New York Times,* October 16, A1, A14.

UCLA Center for Communication Policy. 2001. *The UCLA Internet Report 2001: Surveying the Digital Future.* Los Angeles, CA: UCLA Center for Communication Policy.

UCLA Center for Communication Policy. 2003. *The UCLA Internet Report: Surveying the Digital Future, Year Three.* Los Angeles: UCLA Center for Communication Policy (www.digitalcenter.org/pdf/InternetReportYearThree.pdf).

United Nations. 1945. *United Nations Charter.* New York, NY: United Nation (https://www.un.org/en/about-us/un-charter/full-text).

Vaessen, Martin. 1993. "Evaluation of Population Data: Errors and Deficiencies." Pp. 4-1-4-69 in *Readings in Population Research Methodology, Vol. 1, Basic Tools,* edited by Donald J. Bogue, Eduardo E. Arriaga, and Douglas L. Anderton. Chicago, IL: Social Development Center, for the United Nations Population Fund.

Vaillant, George E. 1995. *The Natural History of Alcoholism Revisited.* Cambridge, MA: Harvard University Press.

van de Vijver, Fons and Kwok Leung. 1997. *Methods and Data Analysis for Cross-Cultural Research.* Thousand Oaks, CA: Sage.

van der Bles, Anne Marthe, Sander van der Linden, Alexandra L. J. Freeman, and David J. Spiegelhalter. 2020. "The Effects of Communicating Uncertainty on Public Trust in Facts and Numbers." *Proceedings of the National Academy of Sciences of the United States of America* 117(14):7672-83.

Van Hoye, Greet and Filip Lievens. 2003. "The Effects of Sexual Orientation on Hirability Ratings: An Experimental Study." *Journal of Business and Psychology* 18:15-30.

Van Kessel, Patrick, Chris Baronavski, Alissa Scheller, and Aaron Smith. 2021. "In Their Own Words, Americans Describe the Struggles and Silver Linings of the COVID-19 Pandemic." *Pew Research Center,* March 5, 2021.

Van Maanen, John. 1982. "Fieldwork on the Beat." Pp. 103–151 in *Varieties of Qualitative Research,* edited by John Van Maanen, James M. Dabbs Jr., and Robert R. Faulkner. Beverly Hills, CA: Sage.

Van Maanen, John. 1995. "An End to Innocence: The Ethnography of Ethnography." Pp. 1–35 in *Representation in Ethnography,* edited by John Van Maanen. Thousand Oaks, CA: Sage.

Vargas, Edward D., Melina Juárez, Gabriel R. Sanchez, and Maria Livaudais. 2019. "Latinos' Connections to Immigrants: How Knowing a Deportee Impacts Latino Health." *Journal of Ethnic and Migration Studies* 45(15):2971-88.

Vargas, Roberto B., Gery W. Ryan, Catherine A. Jackson, Rian Rodriquez, and Harold P. Freeman. 2008. "Characteristics of the Original Patient Navigation Programs to Reduce Disparities in the Diagnosis and Treatment of Breast Cancer." *Cancer* 113:426-33.

Vega, Tanzina. 2014. "To Measure More Diverse America, Solution May Be in Census Questions." *The New York Times,* July 2, A12, A16.

Venkatapuram, Sridhar. 2013. "Subjective Wellbeing: A Primer for Poverty Analysts." *Journal of Poverty and Social Justice* 21:5-17.

Venkatesh, Sudhir. 2008. *Gang Leader for a Day: A Rogue Sociologist Takes to the Streets.* New York, NY: Penguin.

Verba, Sidney and Norman Nie. 1972. *Political Participation: Political Democracy and Social Equality.* New York, NY: Harper & Row.

Verba, Sidney, Norman Nie, and Jae-On Kim. 1978. *Participation and Political Equality: A Seven-Nation Comparison.* New York, NY: Cambridge University Press.

Vernez, Georges, M. Audrey Burnam, Elizabeth A. McGlynn, Sally Trude, and Brian S. Mittman. 1988. *Review of California's Program for the Homeless Mentally Disabled.* Santa Monica, CA: RAND.

Vidich, Arthur J. and Stanford M. Lyman. 2004. "Qualitative Methods: Their History in Sociology and Anthropology." Pp. 27-84 in *The Handbook of Qualitative Research,* 2nd ed., edited by Norman Denzin and Yvonna S. Lincoln. Thousand Oaks, CA: Sage.

Villarreal, Andrés. 2010. "Stratification by Skin Color in Contemporary Mexico." *American Sociological Review* 75:652-78.

Vincus, Amy A., Chris Ringwalt, Melissa S. Harris, and Stephen R. Shamblen. 2010. "A Short-Term, Quasi-Experimental Evaluation of D.A.R.E.'s Revised Elementary School Curriculum." *Journal of Drug Education* 40:37-49.

Viswanathan, Madhu. 2005. *Measurement Error and Research Design.* Thousand Oaks, CA: Sage.

Vogels, Emily A. 2020. "From Virtual Parties to Ordering Food, How Americans Are Using the Internet During COVID-19." *Pew Research Center,* April 30.

Wageman, Ruth. 1995. "Interdependence and Group Effectiveness." *Administrative Science Quarterly* 40:145-80.

Wagner III, William E. 2011. *Using IBM SPSS Statistics for Social Statistics and Research Methods.* 3rd ed. Thousand Oaks, CA: Sage.

Wagner III, William E. 2019. *Using IBM® SPSS® Statistics for Research Methods and Social Science Statistics.* 7th ed. New York, NY: Sage.

Waksberg, Joseph. 1978. "Sampling Methods for Random Digit Dialing." *Journal of the American Statistical Association* 73:40-6.

Walker, Michael E., Stanley Wasserman, and Barry Wellman. 1993. "Statistical Models for Social Support Networks." *Sociological Methods & Research* 22:71-98.

Walker, Robert, Mark Tomlinson, and Glenn Williams. 2010. "The Problem with Poverty: Definition, Measurement and Interpretation." Pp. 353-76 in *The SAGE Handbook of Measurement,* edited by Geoffrey Walford, Eric Tucker, and Madhu Viswanathan. Thousand Oaks, CA: Sage.

Wallace, Walter L. 1983. *Principles of Scientific Sociology.* New York, NY: Aldine.

Wallgren, Anders, Britt Wallgren, Rolf Persson, Ulf Jorner, and Jan-Aage Haaland. 1996. *Graphing Statistics and Data: Creating Better Charts.* Thousand Oaks, CA: Sage.

Walters, Pamela Barnhouse, David R. James, and Holly J. McCammon. 1997. "Citizenship and Public Schools: Accounting for Racial Inequality in Education for the Pre- and Post-Disfranchisement South." *American Sociological Review* 62:34-52.

Wang, Qi, Nolan Edward Phillips, Mario L. Small, and Robert J. Sampson. 2018. "Urban Mobility and Neighborhood Isolation in America's 50 Largest Cities." *PNAS* 115(30):7735-40.

Ward, Lester Frank. 1897. *Dynamic Sociology,* 2nd ed. New York, NY: D. Appleton.

Warner, W. Lloyd, Marchia Meeker, and Kenneth Eells. 1960. *Social Class in America: A Manual of Procedure for the Measurement of Social Status.* New York, NY: Harper & Row.

Warr, Mark. 1995. "The Polls—Poll Trends: Public Opinion on Crime and Punishment." *Public Opinion Quarterly* 59:296-310.

Washington, Harriet A. 2006. *Medical Apartheid: The Dark History of Medical Experimentation on Black Americans from Colonial Times to the Present.* New York: Doubleday.

Watson, Charles G., Curt Tilleskjor, E. A. Hoodecheck-Schow, John Pucel, and Lyle Jacobs. 1984. "Do Alcoholics Give Valid Self-Reports?" *Journal of Studies on Alcohol* 45(4):344-48.

Weatherby, Norman L., Richard Needle, Helen Cesari, Robert Booth, Clyde B. McCoy, John K. Watters, Mark Williams, and Dale D. Chitwood. 1994. "Validity of Self-Reported Drug Use Among Injection Drug Users and Crack Cocaine Users Recruited Through Street Outreach." *Evaluation and Program Planning* 17(4):347-55.

Weaver-Hightower, Marcus B. 2013. "A Mixed Methods Approach for Identifying Influence on Public Policy." *Journal of Mixed Methods Research* 20:1-24.

Webb, Eugene J., Donald T. Campbell, Richard D. Schwartz, and Lee Sechrest. 2000. *Unobtrusive Measures,* rev. ed. Thousand Oaks, CA: Sage.

Weber, Max. 1949. *The Methodology of the Social Sciences.* Translated and edited by Edward A. Shils and Henry A. Finch. New York, NY: Free Press.

Weber, Robert Philip. 1990. *Basic Content Analysis,* 2nd ed. Newbury Park, CA: Sage.

Wechsler, Henry, Jae Eun Lee, Meichun Kuo, Mark Seibring, Toben F. Nelson, and Hang Lee. 2002. "Trends in College Binge Drinking During a Period of Increased Prevention Efforts." *Journal of American College Health* 50:203-17.

Weinberg, Darin. 2000. "'Out There': The Ecology of Addiction in Drug Abuse Treatment Discourse." *Social Problems* 47:606-21.

Weinhardt, Michael. 2020. "Ethical Issues in the Use of Big Data for Social Research." *Historical Social Research* 45(3):342-65.

Weisburd, David, Cynthia M. Lum, and Anthony Petrosino. 2001. "Does Research Design Affect Study Outcomes in Criminal Justice?" *Annals of the American Academy of Political and Social Science* 578:50-70.

Weiss, Carol H. 1993. "Where Politics and Evaluation Research Meet." *Evaluation Practice* 14: 93-106.

Wellman, Barry and Keith Hampton. 1999. "Living Networked in a Wired World." *Comparative Sociology* 28:1-12.

Wellman, Barry and Scot Wortley. 1990. "Different Strokes From Different Folks: Community Ties and Social Support." *American Journal of Sociology* 96:558-88.

Wellman, Barry, Anabel Quan Haase, James Witte, and Keith Hampton. 2001. "Does the Internet Increase, Decrease, or Supplement Social Capital? Social Networks, Participation, and Community Commitment." *American Behavioral Scientist,* 45:436-55.

Wellman, Barry. 2004. "Connecting Communities: On and Offline." *Contexts* 3:22-28.

Wells, William, Michael R. Cavanaugh, Jeffrey A. Bouffard, and Matt R. Nobles. 2012. "Non-Response Bias With a Web-Based Survey of College Students: Differences From a Classroom Survey About Carrying Concealed Handguns." *Journal of Quantitative Criminology* 28:455-76.

Wengraf, Tom. 2001. *Qualitative Research Interviewing: Biographic Narrative and Semi-structured Methods.* Thousand Oaks, CA: Sage.

West, Steven L. and Ken K. O'Neal. 2004. "Project D.A.R.E. Outcome Effectiveness Revisited." *American Journal of Public Health* 94:1027-29.

Wheeler, Peter M. 1995. *Social Security Programs Throughout the World—1995.* Research Report #64, SSA Publication No. 13–11805.

White, Aaron M., Courtney L. Kraus, Lindsey A. McCracken, and H. Scott Swartzwelder. 2003. "Do College Students Drink More Than They Think? Use of a Free-Pour Paradigm to Determine How College Students Define Standard Drinks." *Alcoholism: Clinical and Experimental Research* 27:1750-56.

White, Michael J. 1993. "Measurement of Population Size, Composition, and Distribution." Pp. 1-1-1-29 in *Readings in Population Research Methodology, Vol. 1, Basic Tools,* edited by Donald J. Bogue, Eduardo E. Arriaga, and Douglas L. Anderton. Chicago: Social Development Center, for the United Nations Population Fund.

White, R. M. 2000. "Unraveling the Tuskegee Study of Untreated Syphilis." *Archives of Internal Medicine* 160(5):585-98.

Whyte, William Foote. 1955. *Street Corner Society,* 2nd ed. Chicago, NY: University of Chicago Press.

Whyte, William Foote. 1991. *Social Theory for Social Action: How Individuals and Organizations Learn to Change.* Newbury Park, CA: Sage.

Wiles, Rose, Amanda Coffey, Judy Robinson, and Sue Heath. 2012. "Anonymisation and Visual Images: Issues of Respect, 'Voice' and Protection." *International Journal of Social Research Methodology* 15:41-53.

Wikipedia. 2022. "Historical Polling for United States Presidential Elections." Online. Downloaded 4/13/2022 from (https://en.wikipedia.org/wiki/Historica_polling_for_United_States_presidential_elections?msclkid=e2932996b5611eca8075e3f04f64e7a).

Wilkerson, Isabel. 2020. *Caste: The Origins of Our Discontents.* New York, NY: Random House.

Willer, David and Henry A. Walker. 2007. *Building Experiments: Testing Social Theory.* Stanford, CA: Stanford University Press.

Williams, Kirk R. and Richard Hawkins. 1986. "Perceptual Research on General Deterrence: A Critical Review." *Law and Society Review* 20:545-72.

Wilson, Reid. 2017. "New Report Finds That Voter Turnout in 2016 Topped 2012." *The Hill,* March 16. Retrieved from (http://thehill.com/homenews/statewatch/324206-new-report-finds-thatvoter-turnout-in-2016-topped-2012).

Wilson, William Julius. 1998. "Engaging Publics in Sociological Dialogue Through the Media." *Contemporary Sociology* 27:435-38.

Wines, Michael. 2017. "Accuracy at Risk as Census Shifts to Online Count." *The New York Times,* December 10:1, 20.

Wines, Michael. 2020. "After Virus Delays, Census Must Scramble to Avoid Undercount." *The New York Times, April 18* (https://www.nytimes.com/2020/04/18/us/coronavirus-census.html).

Wines, Michael. 2021. "Despite Many Obstacles, the 2020 Census Gets a Cautious Thumbs-Up." The New York Times, September 14 (https://www.nytimes.com/2021/09/14/us/census-analysis-apportionment.html).

Wines, Michael and Maria Cramer. 2022. "2020 Census Undercounted Hispanic, Black and Native American Residents." The New York Times, March 20.

Witkin, Belle Ruth and James W. Altschuld. 1995. *Planning and Conducting Needs Assessments: A Practical Guide.* Thousand Oaks, CA: Sage.

Witte, James C., Lisa M. Amoroso, and Philip E. N. Howard. 2000. "Research Methodology: Method and Representation in Internet-Based Survey Tools—Mobility, Community, and Cultural Identity in Survey 2000." *Social Science Computer Review* 18:179-95.

Wolcott, Harry F. 1995. *The Art of Fieldwork.* Walnut Creek, CA: Alta Mira Press.

Wolf, Amanda, David Turner, and Kathleen Toms. 2009. "Ethical Perspectives in Program Evaluation." Pp. 170-84 in *The Handbook of Social Research Ethics,* edited by Donna M. Mertens and Pauline E. Ginsberg. Thousand Oaks, CA: Sage.

Woodie, Alex. 2020. "Five Ways to Improve the Accuracy of Polls." *Datanami,* November 2 (https://www.datanami.com/2020/11/02/2020-election-five-ways-to-improve-the-accuracy-of-polls/).

World Health Organization (WHO). 2005. *WHO Multi-Country Study on Women's Health and Domestic Violence Against Women: Summary Report of Initial Results on Prevalence, Health Outcomes and Women's Responses.* Geneva: World Health Organization.

World Health Organization (WHO). 2013. "Management of Substance Abuse" (www.who.int/substance_abuse/facts/alcohol/en).

World Health Organization (WHO). 2015. WHO Issues Best Practices for Naming New Human Infectious Diseases. Retrieved December 21, 2021 from (https://www.who.int/news/item/08-05-2015-who-issues-best-practices-for-naming-new-human-infectious-diseases).

World Health Organization (WHO). n.d. *Process of Translation and Adaptation of Instruments* (www.who.int/substance_abuse/research_tools/translation/en).

Wu, Stephen. 2021. "Leadership Matters: Police Chief Race and Fatal Shootings by Police Officers." *Social Science Quarterly* 102(1):407-19.

Wunsch, Guillaume J. and Marc G. Termote. 1978. *Introduction to Demographic Analysis: Principles and Methods.* New York, NY: Plenum Press.

Xu, Yili, Mora L. Fiedler, and Karl H. Flaming. 2005. "Citizens' Judgment Discovering the Impact of Community Policing: The Broken Windows Thesis, Collective Efficacy, and Citizens' Judgment." *Journal of Research in Crime and Delinquency* 42:147-86.

Yang, Song, Franziska B. Keller, and Lu Zheng. 2017. *Social Network Analysis: Methods and Examples.* Thousand Oaks, CA: Sage.

Young, Robert L. 1992. "Religious Orientation, Race and Support for the Death Penalty." *Journal for the Scientific Study of Religion* 31:76-87.

Zaret, David. 1996. "Petitions and the 'Invention' of Public Opinion in the English Revolution." *American Journal of Sociology* 101:1497-555.

Zhang, Sarah. 2015. "The Pitfalls of Using Google Ngram to Study Language." *Wired,* October 12 (https://www.wired.com/2015/10/pitfalls-of-studying-languagewith-google-ngram/).

Zhong, Juan and Jeffrey J. Arnett. 2014. "Conceptions of Adulthood Among Migrant Women Workers in China." *International Journal of Behavioral Development* 38:255-65.

Zielbauer, Paul. 2000. "2 Cities Lag Far Behind the U.S. in Heeding the Call of the Census." *The New York Times,* April 21, A21.

Zimbardo, Philip G. 2004. "A Situationist Perspective on the Psychology of Evil: Understanding How Good People Are Transformed into Perpetrators." Pp. 21-50 in *The Social Psychology of Good and Evil: Understanding Our Capacity for Kindness and Cruelty,* edited by Arthur G. Miller. New York, NY: Guilford Press.

Zimbardo, Philip. 2007. *The Lucifer Effect: Understanding How Good People Turn Evil.* New York, NY: Random House.

Zitner, Aaron. 1996. "A Cloudy Gaze into a Crystal Ball." *The New York Times,* September 26, D1, D14.

INDEX

Abductive analysis, 404
Abma, Tineke, 474
Abortion, 272, 326, 326e
Academic motivations for social research, 7
Acharya, Arnab, 531
Action research, 473
Advocacy and intervention, 423
African American Health Neighborhood Assessment Scale (AAH NAS), 138, 139e
Aggregate matching, 238–39
Agreement bias, 271
Alcohol Abstinence Self-Efficacy Scale (AASE), 271, 272e
Alcohol use disorder (AUD) and intoxication
  aggression and, 204–206, 204e, 205e
  content validity and, 138
  contextual effect and, 219
  criterion validity and, 138
  difference between reliability and validity in, 140e
  face validity and, 137
  idiographic explanations of, 210
Alone Together: Why We Expect More From Technology and Less From Each Other, 392
Amato, Paul, 140
Ambwani, Suman, 547–49, 553
American Community Survey, 497
American Enterprise Institute, 114
American Journal of Sociology, 40
American Medical Association (AMA), 78
American Psychologist, 82–83
American Sociological Association (ASA), 25, 379
  Code of Ethics, 81, 253, 585
American Sociological Review, 36, 40
Anderson, Elijah, 210, 356, 390, 397, 402
Anomalous findings, 62
Anonymity, 294, 379, 422
Antley, Angus, 91
Applied research reports, 572–77
  California Welcome Baby Program, 600–605, 602–603
Appropriate boundaries, 379
Arnett, Jeffrey, 438–39, 447
ARPANET, 505
Art, qualitative analysis as, 416–418, 417e
Asch, Solomon, 229
Assignment variable, 241
Assisted Living Medicaid Waiver Program, 465
Association
  describing, 341–342, 342e
  evaluating, 343–345, 343–345e
  graphing, 340–341, 340–341e
Asynchronous online interviews, 372–73
"Attitude Toward Blacks" (ATB) scale, 225, 226e, 227e
Authenticity, 66, 70
Availability (convenience) sampling, 175–76
Avoiding harm of research participants, 85–88

Bail, Christopher, 531
Balanced response choices, 132, 133e
Balance point, mean as, 329e, 318
Bar charts, 318, 319e, 597e
Base number N, 311
Basic science, 24
Baumann, Robert, 492
Baumrind, Diana, 82–83, 86
Baxter, Caity, 462
Becker, Howard, 397
Beecher, Henry K., 80
Before-and-after designs, 231, 235–236, 236e, 237e
  "Behavioral Study of Obedience," 82
Behavior coding, 276
Behind the Shock Machine: The Untold Story of the Notorious Milgram Psychology Experiments, 84
Bellah, Robert, 372
Belmont Report, 80
Belousov, Konstantin, 360
Bendix, Reinhard, 523, 536
Beneficence, 80
Berk, Richard A.
  benefits vs. risks of research by, 93
  on contextual effect, 219
  deductive research, 59
  on deterrence, 251
  external validity and, 68
  field experiment, 235, 249
  internal validity and, 70
  literature review, 41, 46
  mixed methods used by, 442
  selective distribution of benefits and, 254
  simple or complex outcomes and, 476
Berk, Richard A., Inter-university Consortium for Political and Social Research (ICPSR) and, 494
Besharov, Douglas, 114
Beta coefficients, 340
Bias
  coverage, 289
  minimizing the risk of, in survey questions, 267
  selection, 245
  social desirability, 269–71

Bibliographic databases, 41
Big data, 505
  background on, 538–541, 540e
  ethical issues in, 544–547, 547e
  examples of research using, 541–543, 541–543e
Bimodal distributions, 317
Binge drinking, 117–18, 123, 135. *See also* Alcohol use disorder (AUD)
Bipolar response options, 265, 266e
Birth cohorts, 206
Bivariate distributions. *See* Cross-tabulation
Black box evaluation, 471, 473
Blogging
  illogical reasoning and, 12
  overgeneralizing and, 10, 12
Boas, Franz, 351
Bohn, Angela, 511
Bond, Robert, 515–16
Boolean search logic, 42
Boot camp programs, 476
Boston McKinney Project, 476, 493–494, 494e
Boundaries, appropriate, 379
Brazilian American Cultural Institute, 119
Brewer, John, 433
British Journal of Criminology, 50
Brooks, Clem, 531
Brown, Ryan, 434
Brown 2013, 440–41, 447
Buchta, Christian, 510
Bureau of Labor Statistics (BLS), 497
Bush, George W., 461
Butler, Dore, 126

Cain, Carole, 210
California Welcome Baby Program, 600–605, 602–603e
Campbell, Donald, 432
Carlisle Indian School, 442, 443e
Case oriented historical research, 524
Case-oriented understanding, 404
Case studies, 355
  designs, comparative, 570–571, 570–571e
  evaluating a public health program, 102–5
  likelihood of voting, 313–314, 314–315e
  patient-provider race concordance and minority health outcomes, 355–356, 356–357e
Causality
  idiographic, 196, 209–10
  mechanism, 218
  nomothetic, 189, 200–201, 201e, 210
  nonspuriousness and, 206–209, 207e
  in quasi-experiments, 242
  randomization and, 208, 208e
  statistical control and, 208–209, 209e
  time order and, 213
  in true experiments, 235, 238
Causal network model, 425, 426e
Causal validity, 66, 69, 248
Cellphones/smartphones
  percentage of households with, 257, 257e
Census Bureau, U. S., 112–114, 113e
  data collected by, 526–527, 526e
  translations of surveys by, 279
Center for Epidemiologic Studies (CES-D), 137, 137e, 269, 270e
Centers for Disease Control and Prevention (CDC), 509
Central Intelligence Agency (CIA), 491
Central tendency, 316, 327–331, 329e, 331e
  distributions differing in variability but not, 331–332, 332e
Certificate of Confidentiality, 92
Ceteris paribus condition, 208
Change, resistance to, 12, 22
Chart Builder, 594–596e, 562
Chi-square, 344, 344e
Civil rights movement, 118
Clark, Andrew, 413
Clinton, Hillary, 162
Closed-ended (fixed-choice) questions, 125, 437
Clusters, 172
Cluster sampling, 165–166, 166e
Cocaine usage, graphs of, 320, 321e
Code of Ethics, ASA, 81, 253, 585
Code of the Street, 356
Coding, 306
  behavior, 276
  content, 128
  design, for variables to be measured, 579–583, 580–581e
Cognitive interviews, 276
Cohen, Susan, 243
Cohort studies, 206
Coin toss, 158
Coleman, James, 432
Collaborative Institutional Training Initiative (CITI), 98
Collective efficacy, 196
Combined and compressed distributions, 325–326, 325–326e
Combined frequency displays, 325–326, 325–326e
Combined questions in indexes, 272–74
Common Rule, 81
Communicating with the public, 584
Community-based participatory research (CBPR), 449, 450e
Community Mental Health Act Amendments, 456
Comparative methods, 531
  case study designs, 570–571, 570–571e
  cautions for, 571–573, 572e
  cross-sectional research, 561–565, 562–566e
  historical research, 567–569, 569e
Comparison groups, 230, 232
  sources of internal invalidity reduced by, 244–45
Comparison of levels of measurement, 135
Compensatory rivalry, 247
Completeness, 178
Complete observers, 358
Complete participants, 358
Complex mixed-methods designs, 473–476, 474–475e
Compressed frequency displays, 325–326, 325–326e
Computer-assisted personal interviewing (CAPI), 287

Computer-assisted qualitative data analysis, 451–454, 451–454e
Computer-assisted telephone interviews (CATI), 286
Computer-mediated communication (CMC), 8
Concepts
  indicators from, 109–115, 117e
  from observations to, 122
  organization, categorization, and condensation in qualitative data analysis, 395, 398
  overlapping dimensions of, 269, 269e
Concurrent validity, 138
Confidence intervals, 178–179, 179e
Confidentiality and privacy, 92, 293, 379, 422–24
Conflicts of interest, 85
Conflict theories, 39
Confusing phrasing in survey questions, 260–262, 263e
Conjunctural historical research, 525
Constants, 118
Constructivist philosophy, 354
Construct validity, 138–39
Contact summary form, 421e, 395
Contamination, 247
Content, coding, 128
Content analysis, 574–575, 576–577e
  design coding, 579–583, 580–581e
  developing appropriate statistical analyses for, 583–584, 585e
  ethical issues in, 555
  identifying a population of documents or other textual sources for, 577–578, 578e
  selecting a sample of units for, 549
Content validity, 138
Context, 219
  effects, 278
  orientation to social, 353
Contingency tables, 337–339, 337–340e
Contingent questions, 267
Continuous measures, 134
Control groups, 213
  before-and-after designs, 231, 235–236, 236e, 237e
  ex post facto, 231, 234–235, 234e
  nonequivalent designs, 238
  quasi-experiments, 238–42
Convenience sampling, 175
Convergent validity, 139
Conversation analysis, 432–433, 434e
Cooper, Kathleen, 8
Core Alcohol and Drug Survey, 135
Correlational analysis, 338–40
Correlation coefficient, 339
Corroboration and legitimization of conclusions, 400
Cost-benefit analysis, 503, 504e
Council of European Social Science Data Archives (CESSDA), 499
Counterfactuals, 200, 201e
Coverage bias, 289
Cover letters, 283–284, 284e
Covert observers, 358, 360
Covert participation, 361–62
Cowan, Gloria, 553
Crabtree, Benjamin, 391
Creative Interviewing, 369
Cress, Daniel, 410
Crime
  committed by illegal immigrants, 194
  community treatment and diversion programs and, 194
  effect of informal social control on, 193e
  effect of social disorder and collective efficacy on rates of, 194e
  graphing association of, 341, 341e
  neighborhood disorder indicators, 446–448, 447e
  nonexperimental designs on, 206, 207e
  rates of homicides, 189e
  recidivism, 476
  self-control and, 197–198, 198e
  structural disadvantage and, 210, 210e
  Transitional Aid Research Project (TARP) and, 464–65
  units of observation and, 197
Crime Control and Safe Streets Act, 481
Criterion validity, 138
Cronbach's alpha, 140, 272
Cross-population generalizability, 59, 60e
  contingency tables, 337–339, 337–340e
Cross-sectional research designs, 200
  comparative, 561–565, 562–566e
  repeated, 203–4
Cross-tabulation, 324
  controlling for a third variable, 345–350, 346–350e
  describing association, 341–342, 342e
  evaluating association, 343–345, 343–345e
  graphing associations and, 340–341, 340–341e
Cyberethnography, 357
Czopp, Alexander, differential attrition and, 246

D'Amico, Elizabeth J., 466
Dannefer, Dale, 437
Data
  big, 505
  Bureau of Labor Statistics (BLS), 497
  census, 526–527, 526e
  cleaning of, 306
  collecting unobtrusive measures, 127
  collection of qualitative, 352
  empirical, 56
  grouped, 323–325, 324e
  Henry A. Murray Research Archive, 491, 499
  Integrated Public Use Microdata Series (IPUMS), 496
  Inter-university Consortium for Political and Social Research (ICPSR), 529–533, 530–532e, 539e, 513
  links between theory and, 48e
  other government sources of, 497
  other sources of, 499
  ownership of, 423
  secondary (*See* Secondary data analysis)
  sources of qualitative, 499
  ungrouped, 322–323e
  using available, 127–28

Data analysis, 302
cross-tabulation, 324
displaying univariate distribution, 307–15
displaying univariate distributions, 307–15
ethical, 335–336, 336e, 341
frequency distributions, 322–326, 322–326e
graphs, 318–321, 319–321e
meta-analysis, 353–356, 356–357e
preparing for, 306–7
qualitative (*See* Qualitative data analysis)
regression analysis, 338–40
summarizing univariate distributions, 315
Data entry, 315–316, 317e
Data Preservation Alliance for the Social Sciences, 491
Data visualization, 596–597, 596e
Debriefing, 89, 254
Deception, 90–91, 252–53
Decisions in research, 91–93e
Decker, Scott, 194
Deductive research, 56, 59
Democratization in Latin America, 568, 569e
Demographic analysis, 573–574, 574e
Demographic bookkeeping equation, 543
Demography, 543
Denzin, Norman, 183, 376, 424
Department of Health and Human Services (DHHS), 81
Department of Housing and Urban Development, 491
Dependent variables, 58
Depression, 269, 270e
Descriptive research, 18, 65
Descriptive statistics, 303. *See also* Association
Design coding procedures, 579–583, 580–581e
Deterrence theory, 60
Detroit Area Studies, 267, 498
Dewey, Thomas E., 161
Dichotomies, 135
Dickson-Swift, Virginia, 380
Differences-in-differences design, 238
Differential attrition, 246
Digital ethnography, 357
Dill, Karen, 547, 549, 553
Dillman, 283
Direction of association, 59
Direct observations, 126
Discrete measures, 132
Discriminant validity, 139
Displaying research, 594–596e, 563
Disproportionate stratified sampling, 169
Distributions, sampling, 176–182, 177e, 179e
Diversity of population, 155–56
Divorce rates, 235–236, 237e
Documentation, 420–422, 421e
Documents for content analysis, identifying populations of, 548
Domestic violence
contextual effect and, 219
incident, offender, and outcome variables by victim-offender relationship, 56e
inductive approach to explaining, 63
Inter-university Consortium for Political and Social Research (ICPSR) data on, 493–99
percentage of suspects with subsequent assault by employment and arrest status, 42e
prevalence of, 28, 29e
second responder programs and repeat family abuse incidents, 44–46, 47e
single-article reviews on, 47–50
when does arrest matter with, 49
Double-barreled questions, 266
Double-blind procedures, 247
Double negatives, 266
Douglas, Jack, 369
Drug Abuse Resistance Education (D.A.R.E.), 487–489, 496, 498e, impact analysis, 501e
Drug Policy Research Center (DPRC), Rand, 479
Duckworth, Kenneth, 549
Duncombe, Jean, 380
Duneier, Mitch, 377, 380, 423, 537
Durkheim, Émile, 8, 490, 522
functionalism and, 39

Ecological fallacy, 190–191, 191e
EconLit, 41
Economic and Social Research Council, U. K., 499
Educational achievement, income and, 351–352, 352e
Effect size, 566
Efficiency analysis, 469
Efficiency of survey research, 262
Ego-based commitments, 13
Eichmann, Adolf, 82
Elaboration analysis, 333
Elder, Keith, 374
Electronic surveys, 279–80
Elements, 152–53
E-mail and social ties, 13
Embedded mixed-methods designs, 438
Emergence, 199
Emic focus, 389
Emmel, Nick, 413
Empirical data, 56
Empirical generalizations, 60
Employment candidates, 220–222, 222e, 243–244, 244e, Project New Hope and, 511, 512e
Encyclopedia Britannica, 585
Engels, Friedrich, 39
Ennett, Susan, 454
Enumeration units, 153
Equal probability of selection method (EPSEM), 165
Erikson, Kai, 355, 362
Errors
ecological fallacy and reductionism, 190–192, 191e
idiosyncratic, 136
measurement, 264
random sampling, 185
in reasoning about the social world, avoiding, 13
sampling, 155, 184, 264

Errors (*Continued*)
in survey research, 264
systematic, 136
Ethics
achievement of valid results and, 81–83
in data analysis, 335–336, 336e, 341
deception, 90–91, 252–53
in evaluation, 514–516, 516e
in experimental research, 252–54
in historical and comparative research and content analysis, 555
honesty and openness in, 83, 85
Milgram's obedience experiment and, 79
mixed methods and, 448
in participant observation, 363
politics, and research reports, 582–86
principles in, 81
protection of research participants and, 85
in qualitative data analysis, 454–458, 455e
in qualitative research, 375–81
in secondary data analysis and Big Data, 544–547, 547e
selective distribution of benefits and, 254
in survey research, 293
Ethnography, 356
digital, 357
video, 424
Etic focus, 389
Ettelt, Stefanie, 479
Eurobarometer Survey Series, 499
European Social Survey (ESS), 531
Evaluability assessment, 461
Evaluation
of association, 343–345, 343–345e
black box, 471, 473
formative, 465
of generalizability, 155
of measures, 143
process, 495–499, 497–498e, 499e
of social research questions, 31–33, 32e
Evaluation research, 21, 455
basics of, 491–492, 491e
black box evaluation or program theory, 505–506, 507e
design decisions, 504–513, 507e, 512e
efficiency analysis, 469–71
ethics in, 514–516, 516e
evaluability assessment, 461
history of, 456
impact analysis, 499–502, 501e
needs assessment, 459–60
process, 495–499, 497–498e, 499e
public health program case study, 105
quantitative and qualitative methods in, 474, 476
questions for, 491–504, 494e, 497–499e, 501–502e
researcher or stakeholder orientation, 473–74
simple or complex outcomes and, 510–511, 512e
Evaluation Review, 456
Event-based designs, 206–7
Events research, historical, 555–557, 556e
Event-structure analysis, 555–557, 556e
Examination and display of relationships in qualitative data analysis, 423–425, 424–425e
Excessive devotion to tradition, 13
Exhaustive choices, 125
Expectancies of experimental staff, 247
Experimental groups, 213, 230, 232
Experiments/experimentation
causality in true, 235
designs and time order, 213
ethical issues in, 252–54
history of, 230
limitations of true, 234–35
quasi-, 231–237, 233e
sources of internal invalidity that require attention during, 245
true, 212, 230–38
validity in, 251
Explanatory research, 16, 17e, 475, 475e
deductive research, 56, 59
inductive research, 62–63
Exploratory research, 19, 64
questions with commitment to inductive reasoning, 352
Ex post facto control group designs, 231, 234–235, 234e
External events and causal validity, 244
External validity, 68, 251
Extraneous variables, 208, 347–348, 347–348e
Exum, M. Lyn, on contextual effect, 219

Facebook, 2, 3, 6, 540–541, 540e
descriptive research about, 18
experiments on its users, 544, 546–547, 547e
synchronous interviewing on, 373
value of, 15e
Face validity, 137
Factorial surveys, 249
Fallon, Kathleen, 536
Family Justice Center (FJC) Initiative, 461
Favorable Attitudes Toward Antisocial Behavior Scale, 127–128, 128e
Fears, Darryl, 119
Feasibility of research, 35
Federal Bureau of Investigations (FBI), 128, 328
Federal Interagency Committee on Education (FICE), 119
Federal Policy for the Protection of Human Subjects, 81
Fedyuk, Olena, 129
Feedback, 458
Fence-sitting, minimization of, 271
Fenno, Richard, 368
Fenton, Steve, 514
Field experiments, 235, 249
Field notes, 387–389, 388e, 420–422, 421e
Field research
developing and maintaining relationships in, 363, 366
entering the field in, 363
ethnographic, 356
managing the personal dimensions in, 367–68

participant observation, 358–63
sampling people and events in, 384–386, 385–386e
taking notes in, 387–389, 388e
Fieldwork, *See also* Participant observation
Filter questions, 262, 263e
Finn, Peter, 462
Fischer, Constance, 404
Fiske, Donald, 432
Fixed-sample panel designs, 197–199, 198e
Flewelling, Robert, 454
Floaters, 267–268, 267e
Focus groups, 276, 358, 374, 379
Food and Drug Administration, 81
Forced-choice questions, 272
Formative evaluation, 465
Fourth generation evaluation, 474
Fowler, James, 515–16
Frank, David John, 528
Franklin, Mark, 533
Freedom of Information Act (FOIA), 555
Frequency distributions, 322–326, 322–326e
Frequency polygons, 318, 320, 320e
Freshman, Audrey, 299e
Fromme, Kim, 466
Functionalism, 39

Gallup polls, 161
Gamma, 330
Geis, Florence, 126
Generalizability, 66, 68
census and, 157
evaluating, 155
in qualitative research, 182, 375
sample, 59, 60e, 248–49
in true experimentation, 248–51
Generalizations, empirical, 60
Generalizing, 11–12
General Social Survey (GSS), 184, 262–63, 491
International Social Survey Program, 491
likelihood of voting and, 313–314, 314–315e
nonresponse to, 294–295, 295e
Gerth, Matthias, 547–49, 553
Ghetto: The Invention of a Place, the History of an Idea, 537
Gill, Hannah, 380
Gilligan, Carol, 409
Ginsberg, Jeremy, 509
Goal-free evaluation, 473
Gobo, Giampietro, 183, 375
Goffman, Erving, 361
Google Flu Trends, 509
Google Scholar, 44
Governmental Accounting Standards Board (GASB), 457, 469
Grand tour questions, 370
Graphs, 318–321, 319–321e
of association, 340–341, 340–341e
Great Depression, the, 161, 456, 526
Griffin, Larry, 554, 555–557, 556e
Grounded theory, 428–430, 429e, 430e
Group-administered surveys, 283
Grouped data, 323–325, 324e
Group randomization, 478
Group units of analysis, 195–96
Guba, Egon, 474
Guesstimates, 128

Habits of the Heart, 372
Hacker, Karen, 418
Hampton, Keith N.
applied research by, 24
qualitative methods used by, 23
Hampton and gupta qualitative methods used by, 23
Handbook of Research Design and Social Measurement, 275
Haphazard sampling, 176
Hard, 584
Harris, David, 114, 555–557, 556e
Harris, Louis, 204
Harvard-MIT Data Center, 491
Harvard Project on Effective Interventions, 475
Hauser, Marc, 85
Hawkins, Donald, 140
Hawthorne effect, 248
Head Start, 239–40
Health and Retirement Study (HRS), 498
Health Insurance Portability and Accountability Act (HIPAA), 92, 99
Health Research Extension Act of 1985, 480
Heckman, James, 468
Henry A. Murray Research Archive, 491, 499
Herek, Gregory, 100
Hermeneutic circle, 370–371, 371e
Hironaka, Ann, 528
Histograms, 318, 319e
Historical events research, 555–557, 556e
Historical process research, 557–558, 558e, 559e
Historical research methods, 522
cautions for, 529
comparative, 567–569, 569e
ethical issues in, 555
events research, 555–557, 556e
overview of, 522
process research, 557–558, 558e, 559e
social science methods, 553–560, 555–556e, 558–559e
Historical social science methods, 553–560, 555–556e, 558–559e
History effect, 244–45
Holistic explanations, 209
Holistic historical research, 524
Homelessness, housing, and mental illness, 422, 444, 449
Homeless persons
Boston McKinney Project and, 476, 493–494, 494e
definition of population, 154
mean total time homeless, 479, 480e
substance use among, 479, 479e
Homicide rates, 189e

Honesty and openness, 83, 85
Hornik, Kurt, 510
Houtzager, Peter, 531
Howard W. Odun Institute for Research in the Social Sciences, 491
Hoyle, Carolyn, 50
Huberman, A. Michael, 395, 398, 422
Human Relations Area Files (HRAF), 534, 565, 566e
Human subjectivity, 353
Humphreys, Laud, 80, 89, 361
Hunt, Dana, 462
Hunter, Albert, 433
Hurricane Katrina, 353
  focus groups, 374
  researcher safety and, 380–81
HyperRESEARCH, 451, 452e
  NVivo, 451–452, 453e, 454e
  stakeholders' stakes, 422, 424e
Hypotheses, 51, 52e

Identical units, 156
Identification of social research questions, 33
Identity disclosure, 293, 378, 423
Idiographic explanations, 196, 209–10, 353
Idiosyncratic errors, 136
Idiosyncratic variation, 272
Illogical reasoning, 12, 22
Immigrants, 561–562, 562–563e
Impact analysis, 499–502, 501e
Inaccurate observation, 10
Incentives, sampling, 179
Income
  education, and voting, 347–350, 348–350e
  growing inequality in, 523–524e
  by highest year of school completed, 351–352, 352e
  interpersonal trust and, 344–345, 344–345e
  voting and, 338–340, 339–340e, 346, 346e
  voting by race and, 350e
Independent variables, 58
Index, 132
Indexes, combined questions in, 272–74
Index of Status Characteristics (I.S.C.), 432
Indicators
  from concepts, 109–115, 117e
  multiple, reliability of, 139–40
Individual matching, 239, 243
Individual randomization, 478
Individual units of analysis, 195
Inductive research, 62–63
Inferential statistics, 185
Informal social control, 193e
Informed consent, 90
Innstrand, Siw Tone, 239
In-person interviews, 293–295, 295e
Inputs, 457
Institute for Survey Research, 105
Institutional review board (IRB), 99, 513
  case study, 102–5
Instrumentation and causal validity, 245
Integrated literature reviews, 49–51
Integrated mixed-methods design, 436
Integrated Public Use Microdata Series (IPUMS), 496
Integrative approaches, 474
Integrity and quality, research, 423
Intelligence Quotient (IQ), 133
Intensive (in-depth) interviewing, 358
Intent-to-treat analysis, 246
Interaction of testing and treatment, 251
Interactive voice response (IVR), 292, 292e
Intercensal percent change, 543
Intercoder reliability, 142
Interitem reliability, 140
Internal validity, 66, 69–70, 249
International data sources, 499
International Social Survey Program (ISSP), 491, 531
International Typographical Union, 435
Internet, the
  activity in one minute, 540e
  communication options on, 454, 455e
  overgeneralization based on, 11–12
  percentage of individuals not using, by personal characteristics, 5, 5e
  percentage of U.S. adults who use, 3–4, 4e
  quantity of data generated on, 540–541, 540e
  surveys via, 296–300, 296e, 299e
Interobserver reliability, 138, 139e
Interpretive comments, 392
Interpretive questions, 276
Interpretivism, 24
Interquartile range, 322
Intersubjective agreement, 24
Inter-university Consortium for Political and Social Research (ICPSR), 525, 529–533, 530–532e, 539e, 513
Interval, sampling, 166
Interval level of measurement, 126e, 128–129, 131e
Interval-ratio level of measurement, 136
Intervening variables, 346–347, 346e
Intervention and advocacy, 423
Interviewing
  in-person, 293–295, 295e
  online, 372
Interview schedules, 273, 279e
Intimate partner violence (IPV)
  defined, 67. *See also* Domestic violence
Intoxication. *See* Alcohol use disorder (AUD) and intoxication
Intraobserver reliability, 139
Intrarater reliability, 139

Jessop, Julie, 380
Job Training Partnership Act (JTPA), 468
John Henry effect, 247
Jones, Jim, 83

Jottings, 367
Journal articles, 598–599, 600e
Journal of Abnormal and Social Psychology, 82
Justice, 80

Kale-Lostuvali, Elif, 411
Kemmis, Stephen, 417
King, Valarie, 140
Kollock, Peter, 19
Kozinets, Robert, 357
Kuhn, Thomas S., 39

Labaw, Patricia J., 276
Labeled unipolar response options, 265, 266e
Labeling theory, 34, 38
Landers, Ann, 9
Landline phones, 257, 257e
Landon, Alfred M., 161
Laub, John, on contextual effect, 219
Leakey, Tricia, 481–82
Ledford, Gerald, 243
Levels of measurement, 125–131, 126e
  comparison of, 135
  dichotomies, 135
  interval, 126e, 128–129, 131e
  nominal, 126–127, 126e, 131e
  ordinal, 126e, 127–128, 128e, 131e
  properties of, 131e
  ratio, 126e, 129–130, 131e
Levinson, Martin, 376, 421
Library of Congress, 491
Lieven, Filip, 250
Likert items, 269
Lincoln, Yvonna, 474
Ling, Rich, quantitative methods used by, 23
Lipset, Seymour Martin, 432, 435
Literal focus, 392
Literary Digest, 161
Literature, research
  integrated reviews, 49–51
  questions to ask about articles in, 592
  searches, 46
  single-article reviews, 47
  systematic reviews, 51–53
  on the web, 44
Longitudinal research designs, 192–193, 195–200, 196e
Lynching, 555–557, 556e

Magazine articles, 41
Mahoney, James, 536
Mailed, self-administered surveys, 282–286, 284e
Mair, Patrick, 510
Majchrzak, Ann, 479
Malinowski, Bronislaw, 351
Manza, Jeff, 531
Margolis, Eric, 411
Markus, M. Lynne, 479
Marshall, Gary, 91
Marx, Karl, 39, 490
Massachusetts Department of Public Health, 103
Matching, 227, 229e
  aggregate, 238–39
  individual, 239, 243
Matrix, 423, 424e
Matrix questions, 278
Maturation and causal validity, 245
McCain, John, 161
McCarter, Susan, 437, 444
McClelland, Charles, 492
McTaggart, Robin, 417
Mean, 328–329, 329e
  median or, 318
Measures/measurement
  of association, 330
  central tendency, 316, 327–331, 329e, 331e
  coding content, 128
  continuous, 134
  defined, 123
  design coding procedures for, 579–583, 580–581e
  dichotomies, 135
  error, 264
  evaluating, 136
  levels of, 125–131, 126e
  making observations for, 126
  photographic, 122–123, 123e
  reliability, 139–40
  unobtrusive, 127
  using available data, 127–28
  validity, 66, 68, 136–39, 142–43
  of variation, 320–41
Mechanical Turk, 249
Mechanism, 218
Median, 317–18
Mediators, 218
MEDLINE, 41, 43, 46
Meta-analysis, 353–356, 356–357e
Method of agreement, 570–571, 570e, 572e
Method of difference, 570–571, 571e, 572e
Michigan Alcoholism Screening Test (MAST), 138
Miles, Matthew, 395, 398, 422
Milgram's obedience experiment/Stanley Milgram, 79, 85–86
  achievement of valid results in, 81–83
  benefits vs. risks of, 92–93
  deception in, 90–91, 252
  diversity of population in, 156
  honesty and openness of, 83, 85
  informed consent in, 94
Mill, John Stuart, 570–571e
Miller, Arthur, 88
Miller, Susan, 394
Miller, William, 391

Mills, C. Wright, 34
Minneapolis Domestic Violence Experiment, 476
See also Berk, Richard A.
Mirowsky, John
phone survey, 286
survey questions, 267
Mixed methods
complex designs, 473–476, 474–475e
embedded designs, 439
ethics and, 448
history of, 432–34
integrated designs, 436
staged designs, 439
strengths and limitations of, 476–481, 477e, 479–480e
types of, 468–476, 469e, 474–475e
Mixed-mode surveys, 291
Mode, 316–17
Moe, Angela, 70, 442
Monitoring the Future, 123
Monotonic patterns, 329
Moore, Barrington, Jr., 523
Moral tales, 406, 408
Morgan, Philip, 526
Morrill, Calvin, 406
Motives for social research, 7
Multidimensional Poverty Index (MPI), 115
Multi-method research, 433
Multiple group before-and-after design, 240
Multiple regression analysis, 353, 353e
Multistage cluster sampling, 166, 166e
Multitrait-multimethod matrix, 432
Mutually exclusive choices, 125
My Lai massacre, Vietnam, 83

Narrative analysis, 435–437, 436e
Narrative explanations, 525
National Archive of Criminal Justice Data, 492
National Archives and Records Administration, 491
National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 80
National Data Program for the Social Sciences, 491
National Digital Preservation Program, 491
National Geographic Society, 289
National Institute of Child Health and Human Development (NICHD), 498
National Institute of Justice, 461
National Institute of Mental Health, 443, 446
National Institute on Aging (NIA), 498
National Institutes of Health (NIH), 92, 98
National Longitudinal Surveys (NLS), 497
National Minority SA/HIV Prevention Initiative Youth Questionnaire, 134
National Opinion Research Center (NORC), 262, 491
National Science Foundation, 491
National Survey of Families and Households, 498
National Technical Information Service (NTIS), 497
National Violence Against Women Survey, 68
Nature, 585
Nazi Germany, 83
Needleman, Carolyn, 435
Needs assessment, 459–60
"Negative Outlook" index, 137, 137e
Negatives, double, 266
Netnography, 357
Neuendorf, Kimberly, 547–49
New England Journal of Medicine, 80
New Jersey Income Maintenance Experiment, 456, 463
Newspaper articles, 41
Newsweek, 310
New York Times, 514, 549
Ngrams, 541–542, 541e
Nixon, Richard M., 266
Nominal level of measurement, 126–127, 126e, 131e
Nomothetic causal explanations, 189, 200–201, 201e
criteria and cautions for, 219
Nonequivalent control group designs, 238
Nonexperimental designs and time order, 206, 207e
Nonprobability sampling methods, 158, 174
availability (convenience) sampling, 175
purposive sampling, 178–79
quota sampling, 170–172, 171e
snowball sampling, 179
Nonrespondents, 160
Nonresponse, 264
to interviews, 294–295, 295e
to mailed, self-administered surveys, 281
to phone surveys, 290, 290e, 291e
Nonspuriousness, 206–209, 207e
Normal distribution, 334–335, 335e
Normal science, 40
Note taking, 387–389, 388e, 420–422, 421e
Nuremberg War Crime Trials, 78
NVivo, 451–452, 453e, 454e

Obama, Barack, 161, 304
O'Brien, Dan, 507
Observation, 7–8, 8e
participant, 358
Office for Protection From Research Risks, 98
Ogburn, William Fielding, 25
Oklahoma City terrorist bombing, 235–236, 237e
Omnibus surveys, 263
Online interviewing, 372
Open-ended questions, 125, 437
Openness and honesty, 84
Operationalization, 117
of concept of race, 118–19
of social network position, 119
Optical illusions, 7e
Oral history, 526
Ordering of questions, 277–279, 279e
Ordinal level of measurement, 126e, 127–128, 128e, 131e

Organic solidarity, 39
Organisation for Economic Co-operation and Development (OECD), 531
Orientations, research, 23–25
Orshansky, Mollie, 114
Outcomes, program, 457
Outliers, 321
Output, 457
Overgeneralization, 8–9, 8e
Overt observation, 360
Overt participation, 360–61
Ownership of data and conclusions, 423

Pagnini, Deanna, 526
Panel studies, 197–199, 198e
Panel Study of Income Dynamics, 246
Paradigms, 39
Partially spurious relationships, 207e
Participant observation, 358
  choosing a role for, 358–63
  continuum, 377, 377e
  developing and maintaining relationships in, 363–64
  entering the field for, 363
  field notes and, 387–389, 388e, 420–422, 421e
  managing the personal dimensions in, 367–68
  sampling of people and events in, 384–386, 385–386e
Participant observers, 360
Participants, protection of research, 85–92
  avoiding deception in research, except in limited circumstances, and, 91, 99
  avoiding harm in, 85–88
  maintaining privacy and confidentiality in, 92
  obtaining informed consent in, 89–90
Participant well-being, 377
Participatory action research, 449–450, 450e, 421, 473
Paternoster, Raymond, 51, 56
Patient-provider race concordance and minority health outcomes, 355–356, 356–357e
Patterson, Orlando, 119
Patton, Michael Quinn, 475
Patulny, Roger, quantitative methods used by, 23
Pearson, Frederic, 492
Pedulla, David S., field experiment, 235
Peer-reviewed articles, 40
Percentages, 338–339, 339e
Perry, Gina, 84, 87
Personal identification numbers (PIN), 289
Personal motivations for social research, 13–14, 14e
Pew Research Center, 4, 262, 496
Phillips, David, 240
Philosophical perspectives, 24
Phone surveys, 286–293, 289–292e
Photo voice, 444, 444e
Piliavin, Irving, 91
Piliavin, Jane Allyn, 91
Placebo effect, 247
Plagiarism, 584–86
Planetary Society, 265
Planning, sample, 152–58
Plessy v. Ferguson, 118
Policy motivations for social research, 7
Policy research, 479
Polls, 158–159, 159e, 161e
  effect of floaters on, 267–268, 267e
  push, 267
Poor measurement, 264
Population
  assessing diversity of, 155–56
  coverage bias, 289
  defining, 154
  inadequate coverage of, 264
  sample components and, 152–54
Population change, 543
Population composition, 573–574, 574e
Population of documents for content analysis, identifying, 548
Population parameter, 185
Population pyramids, 573–574, 574e
Positivism, 24
Posters, research, 605–607, 606e, 608e
Postpositivism, 24
Posttest measures, 232
Poverty
  contextual effect, 219
  motivation for studying immigrants living in, 13, 14e
  relative, 115
  State Children's Health Insurance Program (SCHIP) and, 500, 502e
PowerPoint, 596, 596e
Predictive validity, 138
Pretests, survey, 276
Price, Richard, 247, 254
Primary sampling units, 153
Priming effect, 230
PRISMA guidelines, 44, 46e
Privacy and confidentiality, 92, 293, 379, 422–24
Probability average, 316–17
Probability of selection, 158
Probability sampling methods, 158, 174
  cluster sampling, 165–166, 166e
  compared, 167–168, 167–168e
  simple random sampling, 160–161, 161e
  stratified random sampling, 162–165, 164e, 165e
  systematic random sampling, 162, 163e
Process consent, 377
Process evaluation, 495–499, 497–498e, 499e
Process research, historical, 557–558, 558e, 559e
Program process, 457
Program theory, 505–506, 507e
Progressive focusing, 389
Project New Hope, 511, 512e
Project SPLASH (Smoking Prevention Launch Among Students in Hawaii), 481–82
Proportionate stratified sampling, 168

Proposals, social research, 100
Protection of research participants, 85–92
  avoiding deception in research, except in limited circumstances, and, 91, 99
  avoiding harm in, 85–88
  maintaining privacy and confidentiality in, 92
  obtaining informed consent in, 89–90
  Pseudoscience, 17
  PsycARTICLES, 41
  Psychological Abstracts, 41
  PsycINFO, 41, 46
  Public Agenda, 475
  Public Health Service, U. S., 80
  Purposive sampling, 178–79
  Push polling, 267

Qualitative comparative analysis (QCA), 438–439, 439e
Qualitative data analysis
  abductive analysis, 404
  alternatives in, 428–440, 429–430e, 434e, 436e, 439e
  as an art, 416–418, 417e
  case-oriented understanding, 404
  causal network model, 425, 426e
  compared with quantitative, 392
  computer-assisted, 451–454, 451–454e
  conversation analysis, 432–433, 434e
  corroboration and legitimization of conclusions in, 400
  documentation, 420–422, 421e
  ethics in, 454–458, 455e
  examination and display of relationships in, 423–425, 424–425e
  features of, 415–419, 417e
  grounded theory, 428–430, 429e, 430e
  narrative analysis, 435–437, 436e
  organization, categorization, and condensation in, 395, 398
  participatory action research, 449–450, 450e, 421
  qualitative comparative analysis, 438–439, 439e
  reflection on researcher's role in, 402
  systematic observation, 446–448, 447–448e
  techniques of, 419–428, 420–421e, 424–426e
  visual sociology, 414
Qualitative Data Archiving Resource Center, 500
Qualitative data sources, 499
Qualitative research, 23–24
  acceptance in, 354–55
  appropriate boundaries, 379–80
  case study, 355
  combining methods of, 410
  data collection, 351
  digital ethnography, 357
  ethical issues in, 381
  ethnography, 356
  evaluation research, 474, 476
  exploratory research, 19, 64, 351–52
  features of, 352
  focus groups, 276, 358, 374, 379
  focus on human subjectivity, 353
  focus on previously unstudied processes and unanticipated phenomena, 352
  fundamentals of, 351
  generalizability in, 182, 375
  history of, 351
  idiographic explanations, 196, 209–10, 353
  interviewing online, 372
  orientation to social context, 353
  participant observation, 358
  researcher safety in, 380–81
  sensitivity to subjective role of researcher in, 368–369, 369e
  See also Explanatory research
Quality
  sample, 179, 181
  secondary data, 501
Qualtrics, 290
Quantitative analysis. *See* Data analysis
Quantitative research, 23–24
  evaluation research, 474, 476
  See also Data analysis
Quartiles, 322
Quasi-experiments, 231–237, 233e
  causality in, 242
Questionnaires
  design of, 275–79
  made attractive, 278
Questions, social research, 33–37
  asking and recording answers to, 371–72
  evaluation research, 491–504, 494e, 497–499e, 501–502e
  grand tour, 370
  literature searches and, 41
  survey (*See* Survey questions)
Quota sampling, 170–172, 171e, 172e

Race
  "Attitude Toward Blacks" (ATB) scale and, 225, 226e, 227e
  lynching and, 555–557, 556e
  operationalizing the concept of, 118–19
  out-of-wedlock births and, 526
  patient-provider race concordance and minority health outcomes, 355–356, 356–357e
  by region of the U.S., 340–341, 340–341e
  voting by income and, 350e
  wedding party photos and, 442, 443e
Randall, Ann, 373
Random digit dialing, 165
Random errors, 136
Randomization/random assignment, 208, 208e, 226–228, 227–228e, 229e
  group, 477, 479
  sources of internal validity reduced by, 245
Randomized comparative change design, 232
Randomized comparative posttest design, 232
Random number tables, 165
Random sampling, 156–157, 228e
  error, 184
  simple, 164

stratified, 162–165, 164e, 165e, 172e
systematic, 162, 163e
Range, 321
interquartile, 322
Rapport and control in in-person interviews, 286–87
Ratio level of measurement, 126e, 129–130, 131e
Rational choice theory, 33e, 38
Raudenbush, Stephen
causal conclusions by, 212
cross-sectional designs, 200–201
nonexperimental designs, 213, 216
systematic observation by, 446–448, 447e
Reactive effects, 360
Reasoning, 12, 22
Recording answers to questions, 372
Reductionism, 198–99
Reevaluating, 12
Refereed journals, 40
Refining of social research questions, 34
Reflexive comments, 392
Reflexivity, 353, 401
Regression and causal validity, 245
Regression-discontinuity design, 241
Regression effect, 245
Reisman, David, 8
Reiss, Albert, 126
Relative poverty, 115
Reliability, measurement, 139
multiple indicators, 139–40
multiple observers, 138, 139e
multiple times, 139
ways to improve, 142–44
Repeated cross-sectional designs, 203–4
Replacement sampling, 165
Reporting
plagiarism and, 584
Reporting research, 597–607, 597e, 600e, 602–603e, 606e
applied research reports, 572–77
on California Welcome Baby Program, 600–605, 602–603e
in journal articles, 598–599, 600e
quantitative and qualitative research, 581
in research posters, 605–607, 606e, 608e
Representative samples, 154, 155e
Request for proposals (RFP), 34
Research circle, 50, 51e
Research designs
comparing, 211–212, 211e
comparison of survey, 301–303, 302e
cross-sectional and longitudinal, 192–200, 196e
evaluation, 504–513, 507e, 512e
mixed methods (*See* Mixed methods)
quasi-experimental, 238–42
units of analysis, 195–99
See also Experiments/experimentation
Researcher or stakeholder orientation in evaluation research, 473–74
Researcher safety, 380–81
Research literature. *See* Literature, research
Research participants. *See* Protection of research participants
Research posters, 605–607, 606e, 608e
Research spiral, 54, 55e
Resistance to change, 12, 22
Respect for persons, 80
Respondent-driven sampling, 173–174, 174e, 175e
Responsive evaluation, 473
Results of research
displaying, 594–596e, 563
reporting, 597–607, 597e, 600e, 602–603e, 606e
writing, 561
Reuter, Peter, 479
Reverse outlining, 561
Reviewing of literature research, 43
Riessman, Catherine Kohler, 210
Ringwalt, Christopher, 454, 463
Risk Skills Training Program (RSTP), 466
Robert Wood Johnson Foundation, 455
Roper Center, 491
Ross, Catherine
phone survey, 286
survey questions, 267
Rossi, Peter, 465, 478
Rubin, Herbert, 178
Rubin, Irene, 178

Sacks, Stanley, 470
Safe and Drug-Free Schools and Communities Act, 457
Safety, researcher, 380–81
Salmons, Janet, 379
Sample(s)
quality of, 179, 181
representative, 154, 155e
of units selected from the population for content analysis, 549
Sample components, 149–151, 150e
Sample generalizability, 59, 60e, 155, 248–49
Sample planning, 152
Sample size, 180–181e, 186–87
Sample statistic, 185
Sampling
frame, 152
methods (*See* Sampling methods)
of people and events in field research, 384–386, 385–386e
purpose of, 152
random, 156–157, 228e
theoretical, 385–386, 386e
units of, 153
Sampling distributions, 176–182, 177e, 179e
Sampling error, 155, 265
estimating, 184, 186
Sampling interval, 166
Sampling methods, 182
nonprobability, 174
probability, 158, 174

Sampson, Robert, 196
causal conclusions by, 212
on contextual effect, 219
cross-sectional design, 200–201
fixed-sample panel design, 204
nonexperimental designs, 213, 216
systematic observation by, 446–448, 447e
Sanders, Andrew, 50
Saturation, 178
Saturation point, 392, 393e
Scales, 274
Schofer, Evan, 528
Schofield, Janet Ward, 375
Scholarly journals, 40
School cohorts, 206
Schorr, Lisbeth, 475
Schreer, George, 91
Schutt, Russell, 126, 422, 437, 445
Science, 14–21
basic, 24
defined, 14
motives for social research, 7
normal, 40
pseudoscience or, 17
types of social research and, 18–21
Scientific approach, 16
Scientific paradigms, 39
Scientific relevance of social research, 36
Seaman, Claire
basic science used by, 24
quantitative methods used by, 23
Seattle–Denver Income Maintenance Experiment, 456
Secondary data analysis, 490
challenges for, 500–504
ethical issues in, 544–547, 547e
sources of data for, 490–500
Secondary sampling units, 153
Second responder programs and repeat family abuse incidents, 44–46, 47e
Selection bias, 245
Selective distribution of benefits, 254
Selective observation, 7, 8e
Self-administered, mailed surveys, 282–286, 284e
Self-control and crime, 197–198, 198e
Self-reports, inadequacy of, 124, 124e
Seniority cohorts, 206
Serendipitous findings, 62
Setting, survey, 279
Sexual abuse/assault, prevalence of, 28, 29e
Sherman, Lawrence
on contextual effect, 219
deductive research, 59
on deterrence, 251
external validity and, 68
field experiment, 235, 249
Inter-university Consortium for Political and Social Research (ICPSR) and, 494
literature reviews, 41, 46
mixed methods used by, 442
selective distribution of benefits and, 254
simple or complex outcomes and, 476
Siegert, Gabriele, 547–49, 553
Simple or complex outcomes, 510–511, 512e
Simple random sampling, 160–161, 161e
Simpson, Brett, 252
Sin, Chih Hoong, 449
Single-article reviews, 47
Skewness, 307
Skip patterns, 262, 263e
Skocpol, Theda, 537, 540
Slater, Mel, 91
Sloboda, Zili, 455, 477, 481
Smith, Marc, 19
Smith, Saundra, 91
Snow, David, 410
Snowball sampling, 179
Social Care Institute for Excellence, 457
Social class in America, 432
Social control, 109e
effect of informal, 193e
Social desirability bias, 269–71
Social health, 112
Social importance of research, 35
Socially desirable behavior, 124, 124e
Social movement organizations (SMO), homeless, 438–439, 439e
Social networking/social media
and changes in size of social networks, 6e
quantity of data generated on, 540–541, 540e
Social Origins of Dictatorship and Democracy, 523
Social research
alternative research orientations for, 23–25
motivations for, 17
questions, 33–36
strategies, 56
strengths and limitations of, 22–23
types of, 18–21
Social research proposals, 100–105
Social research standards, 66–70
Social Science Citation Index (SSCI), 41
Social sciences
approaches in evaluation research, 473
comparative methods, 531
definition of, 16
historical methods, 553–560, 555–556e, 558–559e
Social Security Administration, 499
Social status, 117
Social stratification, 397
Social support, 117
Social theories, 34, 39
links between data and, 48e
and predictions about effect of arrest for domestic assault, 49e
social status in, 117

Social ties
e-mail and, 13
explanatory research on Internet effect on, 16, 17e
Social world, the
avoiding errors in reasoning about, 13
learning about, 4–8
SocINDEX, 41, 46, 495
Socioeconomic status, 123
Sociological Abstracts, 37, 40, 41e
Solomon four-group design, 228, 245e
Specification, 348–350, 349–350e
Split-ballot decision, 263
Split-half reliability, 140
Spurious relationships, 207–208, 207e
use of statistical control to reduce, 209e
Staged mixed-methods designs, 438, 440
Stake, Robert, 395
Stakeholders, 458
approaches in evaluation research, 473
Stald, Gitte, quantitative methods used by, 23
Standard deviation, 333–335, 335e
Standardization of population numbers, 543
Standards, social research, 66–69
Stanford Research Institute, 505
State Children's Health Insurance Program (SCHIP), 500, 502e
States and Social Revolutions, 541
Statistical control, 208–209, 209e
Statistical Package for the Social Sciences (SPSS), 307
Statistical power analysis, 186
Statistical significance, 332
Statistics, 313–314, 314–315e
cross-tabulating variables, 324
displaying univariate distributions, 307–15
ethical data analysis, 335–336, 336e
inferential, 185
measures of central tendency, 316, 327–331, 329e, 331e
meta-analysis, 353–356, 356–357e
regression analysis, 338–40
summarizing univariate distributions, 315–41
See also Association
Stephen, Jason, 585
St. Jean, Peter, 448, 448e
Stokoe, Elizabeth, 406, 411
St. Pierre, Robert, 478
Strategies, social research, 56
See also Explanatory research
Stratified random sampling, 162–165, 164e, 165e, 172e
Strauss, Jaine, 547–49, 553
Street Corner Society, 363
Street-level bureaucrats, 501
Strengths and limitations of social research, 22
Stress among individuals with intellectual disabilities, 233, 233e
Structural disadvantage and crime, 210, 210e
Structure of Scientific Revolutions, The, 39
Subject descriptors, 42
Subject fatigue, 206
Subjective role of researchers in qualitative research, 368–369, 369e
Subjective well-being, 116
Subject well-being, 377
Substance use
applied reports, 605, 606e
cocaine usage graphs, 320, 321e
homelessness and, 479, 479e
self-report and observer ratings of, 477–478, 477e
Subtables, 333
Suicides
out-of-wedlock births and, 526
TV soap operas and, 235, 236e
Survey Monkey, 299–300, 299e
Survey organization, 279–93
cover letters, 283–284, 284e
group-administered surveys, 283
in-person interviews, 293–295, 295e
mailed, self-administered surveys, 282–286, 284e
manner of administration, 279–80
mixed-mode surveys, 291
questionnaire structure, 280
setting, 280
telephone surveys, 286–293, 289–292e
web surveys, 296–300, 296e, 299e
Survey pretests, 276
Survey questions
avoiding confusing phrasing in, 260–262, 263e
avoiding making either disagreement or agreement disagreeable in, 271
combined in indexes, 274
interpretive, 276
maximizing the utility of response categories in, 267
minimizing fence-sitting and floating in, 267–268, 267e
minimizing the risk of bias in, 267
ordering of, 277–279, 279e
refining and testing, 275
Survey research
attractions of, 262
avoiding confusing phrasing in, 265
comparison of designs for, 301–303, 302e
designing questionnaires for, 275
errors in, 264
ethical issues in, 293–94
omnibus, 263
organizing, 279
questionnaires made attractive, 278
translation of questionnaires in, 278
writing questions for, 272
Swiss, Liam, 536
Symbolic interaction theory, 38
Synchronous online interviews, 372–73
Systematic bias, 160
Systematic errors, 136
Systematic literature reviews, 51–53
Systematic observation, 446–448, 447–448e
Systematic random sampling, 162, 163e

Tacit knowledge, 401
Take Charge of Your Life (TCYL), 477
Tavory, Iddo, 404
Tearoom Trade, 80, 89
Teenage health in schools, 114–115, 115e
Telephone surveys, 286–293, 289–292e
Temporal historical research, 524
Ten Point Coalition, 475
Theoretical sampling, 385–386, 386e
Theory
  links between data and, 48e
  of moral development, 409
  program, 505–506, 507e
Theory-driven evaluation, 472
Therapeutic communities (TCs), 470
Thick description, 355
Thill, Kathryn, 547, 549, 553
Thoennes, Nancy, 67
Thomas, Kirsten, 91
Thorne, Barrie, on reflexivity, 401
Time order, 193–195, 198, 206, 207e
Time series designs, 240
Timmermans, Stefan, 404
Tjaden, Patricia, 67
Toepoel, Vera, 579–80
Townsend, MEg, 462
Tradition, excessive devotion to, 13
Transitional Aid Research Project (TARP), 496, 498, 499e
Trend studies, 203–4
Triangulation, 129
Trow, Martin, 432
True experiments, 212, 230–38
  causality in, 235, 238
  generalizability in, 248–51
  limitations of, 234–35
  See also Experiments/experimentation
Truman, Harry S., 161
Trump, Donald, 162
Turkle, Sherry, 392

Unbalanced response choices, 132, 133e
Ungrouped data, 322–323e
Uniform Crime Reports (UCR), 128, 328, 529
Unimodal distributions, 317
Union democracy, 432
United Nations, 491
Units of analysis, 195–99
  ecological fallacy and reductionism, 190–192, 191e
Units of analysis in content analysis, 548
Units of observation, 197
Univariate distributions
  displaying, 307–15
  frequency distributions, 322–326, 322–326e
  graphs, 318–321, 319–321e
  summarizing, 315–41
Unlabeled unipolar response options, 265, 266e
Unobtrusive measures, 127
  available data and, 127–28
  coding content and, 128
Unrepresentative samples, 154, 155e
USA Today, 549
Use and misuse of results, 424
Utility of response categories, 267, 269
Utilization-focused evaluation, 473

Validity, 66
  causal, 66, 69, 248
  construct, 138–39
  content, 138
  criterion, 138
  in experiments, 251
  external, 68, 251
  face, 137
  generalizability, 66, 68, 248–51
  internal, 66, 69, 248
  measurement, 66, 68, 136
  ways to improve, 142–43
Values role in research, 25
Van Hoye, Greet, 250
Van Maanen, John, 364
Variability, 307
  distributions differing not in central tendency, but in, 331–332, 332e
Variable-oriented research, 531
Variables, 117
  concepts, indicators, and, 117e
  vs. constants, 118
  cross-tabulating (*See* Cross-tabulation)
  dependent, 59
  design coding procedures for, 549–53
  extraneous, 208, 347–348, 347–348e
  independent, 58
  intervening, 346–347, 346e
  specification, 348–350, 349–350e
Variance, 333, 334e
Variation, measures of, 320–41
Venkatesh, Sudhir, 378
Versatility of survey research, 262
Victory, Nancy, 8
Video ethnography, 424
Virtual ethnography, 357
Visual sociology, 414
Viterna, Jocelyn, 536
Voluntary participation, 377
Voting
  gender, interpersonal trust, and, 342, 342e
  household income and, 338–340, 339–340e, 346, 346e
  immigrants and, 561–563, 562–563e
  income and education and, 347–350, 348–350e
  international turnout, 561–565, 562e, 564–565e
  likelihood of, 313–314, 314–315e

Walker, Henry A., 254
Ward, Lester Frank, 25
Ward Schofield, Janet, 183
Ware, Norma, 367
Weaver-Hightower, Marcus, 434, 447
Webb, Eugene, 127
Weber, Max, 522, 536
Web surveys, 296–300, 296e, 299e
Weinberg, Darin, 122
Well-being, subject, 377–78
Wellman, Barry, applied research by, 24
Wertz, Frederick, 404
WEXTOR, 249
Whyte, William Foote, 368, 415
  subject well-being and, 377
Wikipedia, 585
Willer, David, 254
Wilson, William Julius, 584
Winship, Chris, 507
Work and Authority in Industry, 523
World Health Organization (WHO), 279
World Income Inequality Database, 499
World War II, 456
Writing research, 560

Yankelovich, Daniel, 475
Young adults
  migration in China, 438

Zaslow, Martha, 471
Zhong, Juan, 438–39, 447
Zimbardo prison simulation study/Philip Zimbardo, 87, 252